

More worries about pollution

Recent events notwithstanding, significant problems lie ahead from increasing concentrations of some constituents of the lower atmosphere. Solutions depend not only on regulation but also on changes in the way environmental science is supported.

THERE is a gas in the troposphere — the lowest 10 km or so of the atmosphere — that is worth singling out. Its background levels in the Northern Hemisphere have doubled over this century, and are now so high that fluctuations frequently take concentrations over health and other damage thresholds. Much of its chemistry is quite well understood, but there are significant uncertainties about its behaviour — its distribution around the globe and associated long-term trends, and factors underlying them. What is more, it neatly exemplifies a problem in the conduct of environmental research associated with developing countries.

The gas in question, ozone, is more famous for its diminishing presence above the troposphere. In fact the global ozone cycle includes transport between the stratosphere and troposphere, which can occasionally give rise to sudden increases at ground level. But the primary tropospheric processes are chemical and photochemical, involving sunlight as well as ozone's primary precursors: nitrogen oxides and volatile organic compounds. As in the stratosphere, ozone itself is not the villain of the piece. In the lower atmosphere, however, problems arise not from chlorofluorocarbon refrigerants but from road transport, industrial solvents and paints, increasing use of fertilizers with population growth, and — in the tropics — biomass burning.

Episodes in which ozone is temporarily enhanced — photochemical smog — have long been notorious in Los Angeles, where reactants are trapped by the topography. Similar immobile pollution arises in Athens, for example. In windier regions to the north, reactions are instigated in cities but, due to their slowness, typically bring about enhanced ozone levels in their host airmasses as they travel well away from the source. That brings about high ozone concentrations in rural areas which, increasingly, are damaging crops. And, whether in cities or downwind of them, periods of human exposure to levels above those readily tolerated by the World Health Organisation are growing both in duration and frequency, giving rise to increased bronchial illness.

A meeting two weeks ago of environment ministers from north-west Europe was the first time that tropospheric ozone received concerted attention at such a high political level. For them, episodic enhancements are a problem of both atmospheric transport and scientific collaboration across national boundaries. In the event they produced only words: requests that the United Nations and the European Union make rapid progress towards abatement protocols, and agreement to cooperate in exchanging data for forecasts. Their difficulty is that abatement involves not only technological fixes to vehicle emissions, which are relatively straightforward, but also, and much more politically challenging, restrictions on the use of road transport. Only when damage to health and national wealth climbs significantly higher, it seems, are citizens or politicians going to grasp the latter nettle.

To reduce the background concentrations is a hemispheric problem. The United Nations and the European Union are developing emission protocols that should help, provided they have teeth, but there is a scientific issue as well. Ozone contributes to greenhouse warming both directly and through its involvement in the chemistry of one of the most climatically active gases, methane. All the more reason, therefore, that we should know much more than we do about its distribution, particularly in the

tropics. To achieve that requires more measurements of ozone and other chemicals. But funding agencies in developed countries will not readily support observing stations in developing ones. At the same time, attempts by distinguished scientists via the UN Environmental Programme (UNEP) to get developing countries and others to stump up, say, US\$5million a year for 25 new monitoring stations have failed to arouse interest among third-world governments.

That is another example of a problem already discussed in these columns (see *Nature* 380, 467; 1996): how to overcome institutional and political barriers to fund well distributed measurements of important geophysical and geochemical quantities — a problem, incidentally, that also undermines the ability of the Intergovernmental Panel on Climate Change to be conclusive (see news, page 455). With tropospheric ozone, the right coalition still needs to be found, within UNEP or elsewhere, to bring home the climatic, health and agricultural significance of our ignorance. But found it surely must be. □

Watch this boom

California's links between science, technology and the economy of the state still set the standard for others to follow.

THE Californian economy is roaring ahead, following a five-year relapse which obliged the Golden State — briefly and uncharacteristically — to question its own unique strengths. As a leading economist told a Californian "science and technology summit" last week, the revival is built on science and its application. The new boom is led by information technology, biotechnology and the movie industry. The awesome strength of the first two in the state (see news, page 459) springs from the virtuous cycle of world-beating science, innovative engineering, and access to venture capital.

The significant role of the federal government in all of this is too often forgotten. Being the largest state in the Union — with the largest congressional delegation and a pivotal role in presidential elections — has helped California secure an unequal share of federal R&D. It has one-eighth of the US population, but gets one-third of NASA's money and a quarter of the Department of Defense's huge research pot. Interestingly, California does less well (fair shares only) when research funds are properly peer-reviewed for quality, as at the National Institutes of Health and the National Science Foundation. But it would be churlish to credit federal generosity for creating California's success in applying science and technology: it has merely served to lubricate the energy, imagination and flexibility which characterizes the place.

In the depths of the recession there was some rumbling for a proper "science and technology policy" for California, but that has quietened. All the state needs now is more of the same. Silicon Valley has already engulfed most of the San Francisco Bay area. Its approach will be replicated elsewhere in the state much more readily than it will be in the rest of the world, which should continue, nonetheless, to measure economic success against the model that California provides. □



Creutzfeldt-Jakob researchers seek greater access to data on cases

Paris. Revelations last weekend in the British press that the United Kingdom has identified five new suspect cases of a novel variant of Creutzfeldt-Jacob Disease (CJD) have brought to the boil simmering tension among researchers over when data concerning new suspect cases should be released, either to the public or to other researchers.

The appearance of the new CJD variant, first identified in March, is widely considered as being possibly related to the consumption of beef contaminated with the agent that causes bovine spongiform encephalopathy. The new suspect cases follow 11 that had previously been confirmed in the United Kingdom (see *The Lancet*, 347, 921; 1996).

The European CJD surveillance network, an informal forum of the national surveillance networks, agreed at a meeting in Edinburgh last month that information on new cases of the novel variant should initially be made public once a month, and perhaps later on a three-monthly basis, if it turns out that this makes more sense in terms of the number of cases being declared.

This approach is endorsed by Robert Will, the head of the CJD Surveillance Unit in Edinburgh, which is supported by the UK Department of Health. "Our intention is to release all the information on a regular basis," he says. "That is how it should be."

But Will also warns that making data available on suspect cases would be "misleading", arguing that many suspect cases turn out not to have CJD at all, much less the novel variant. "It is meaningless to talk about suspect cases," he says, adding that long before the current events, it had been his unit's policy not to disclose details of suspect cases. "My concern is to give accurate information," he says. "There is a lot of speculation about what is happening, and a lot of it is inaccurate."

Will points out, for example, that the original paper on the novel variant of CJD referred to one case with all the symptoms of the variant — the patient died in 1980 at the age of 16 years from CJD-like symptoms — but that this diagnosis was eventually discounted. "I'm reluctant to disclose suspect cases," he says. "If I start saying we have X 'suspect cases', people will say there are X 'cases', and that is very misleading."

Other researchers in the field say that they empathize with Will's concerns about the dangers of misleading the public. But several disagree with the insistence of the Edinburgh surveillance unit not to divulge information on suspect cases. "It's a pity that

these five cases appeared in the *Sunday Times* before I had been informed," says one member of the European CJD network.

Other members of the network agree that there is little point in making public all suspect cases either of CJD, or of its novel variant. Nevertheless they argue that information on 'highly-suspect' cases — for which additional neurological information is available — should be circulated promptly to all researchers working in the field, both in Britain and abroad.

"The United Kingdom has probably already done cerebral biopsies on these five suspect cases," says one European scientist. "But we are still completely in the dark [as to the results]; we still don't know whether these five cases are confirmed or not."

These researchers emphasize the importance of access to new information, even on

suspect cases. "If we are going to have an idea of the evolution of the new variant, we need to know its incidence, and the exact dates when cases occurred," says one.

Other researchers point out that science is "a two-way process." France, for example, has had one confirmed case of the new variant of CJD, but has two other highly-suspect cases. French scientists point out that this information has been made available to UK scientists. "They need to know figures from other European countries, especially at this phase of uncertainty," says one CJD researcher.

"We are going to have to hammer out a definition of what is a highly suspect case," says another, adding that agreement is needed that information on such cases will be released in the same way as that on confirmed cases.

Declan Butler

ESA wants compulsory Earth missions

Munich. Nine teams of European scientists last week presented proposals in a competition organized by the European Space Agency (ESA) to select a new family of scientific Earth observation missions — called Earth Explorers — to fly in the first decade of the next century. ESA's 15 member states will choose three or four missions for further study towards the end of the year.

One candidate mission would measure the Earth's magnetic fields at various altitudes and study their influence on the solar wind. Another would measure precisely the Earth's shape, which in turn would allow a better determination of the net ocean circulation. Remote sensing technology would also be used by a further proposed mission to map glaciers, ice sheets and sea ice, yielding data for research into climate change.

ESA's Earth observation programme so far has consisted mainly of operational satellites geared towards commercial users, such as meteorological services, although the stable includes a series of multi-instrumented remote-sensing satellites, ERS-1 and ERS-2, and the planned Envisat satellite. More than half the data from ERS missions go free of charge to over 270 selected research teams (see *Nature* 374, 665; 1995), on the basis of peer-reviewed competition.

But the idea behind the dedicated family of Earth Explorers is for researchers to be more directly involved in designing instruments to suit their scientific requirements. This 'bottom-up' approach is modelled on

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Rough seas: exaggerated sea-levels as recorded by Europe's ERS-1 satellite.

that used in ESA's successful space science programme (see Briefing, page 461).

To reinforce the Earth Explorer programme, ESA also wants its funding to be compulsory for member states, again in line with the space science programme, which is the only mandatory ESA programme.

Providing the Earth Explorer programme with the same financial status as the science programme would ensure guaranteed levels of funding. ESA will put its request to the space ministers of its member states at a meeting in Brussels next year. Given the tight financial situation which led last year to the capping of the space science programme, however, ESA may find it difficult to win immediate approval for its proposals.

Alison Abbott

DLR, ESA/SPL

Fusion programme 'could aid terrorists'

Washington. A former US nuclear weapons designer has criticized a joint programme between the United States and Russia aimed at exploring new ways of producing fusion energy, claiming that it could produce information that terrorists could use to develop a miniature neutron bomb. Anti-nuclear organizations also say that the techniques being developed undermine non-proliferation efforts.

The programme between the Los Alamos National Laboratory (LANL) in New Mexico and its Russian counterpart, Arzamas 16, is part of the US Department of Energy's 'lab-to-lab' programme. This is intended to keep Russian weapons scientists gainfully employed, and to encourage them not to sell their services to countries bent on acquiring nuclear weapons.

The two laboratories have already carried out 18 joint experiments in which chemical explosives are used to create a powerful electrical pulse. This is converted into a magnetic field, which is then used to compress a pellet of deuterium and tritium, triggering thermonuclear fusion. Andrei Sakharov, the renowned Soviet weapons scientist and dissident, first proposed this approach in the 1950s.

But writing in the *Wall Street Journal* last month, Sam Cohen, a former weapons designer at LANL, claimed that while the experiments are a long way from creating a new source of fusion energy, they are creating a device with "all the hallmarks of a 'pure fusion' neutron bomb" that could prove an ideal "briefcase bomb" for terrorists.

His thesis is that although the tiny amounts of fusion energy released in the experiments are insufficient for a fusion bomb, the devices that are being tested successfully burn deuterium and tritium to produce "deadly high-energy neutrons". This burning process is identical, he says, to that used in conventional neutron bombs. Cohen, who invented the neutron bomb, also claims that the US government abandoned similar research in the 1950s

after he had raised similar objections.

Cohen's warning has met with a mixed response. "Pretty darned far-fetched," says Paul White, programme manager in LANL's centre for international security affairs. "There is no practical way that any scientist could imagine of converting this into a bomb," he says, adding that even if Cohen's thesis is correct, any device would be too big to be useful as a weapon. Making a weapon from such a pulsed energy device would be comparable to making one from a tokamak fusion reactor, he adds.

But Cohen argues that documents from the programme show that it is expected soon to produce as much as 200 pounds of TNT-equivalent fusion yield, and would release large quantities of lethal high-energy neutrons. Los Alamos National Laboratory confirms that an experiment of this yield is



possible, but says that it will require the use of two of the one-ton devices (called generators), around 1,000 pounds of explosives and a one-ton capacitor bank. Hardly a deliverable weapon, White says, claiming that further miniaturization of the apparatus is unlikely, given that Russia has been optimizing such techniques for 40 years.

But Cohen points out that early nuclear devices were also bulky, whereas modern warheads can be the "size of a soccer ball". Moreover, he says that the experiments carried out seem to have overcome the technical hurdle that stymied years of earlier efforts by US weapons designers to develop a pure fusion weapon, namely creating an explosive nuclear device that does not require a fission trigger.

The latter claim is questioned by Stephen Dean, president of Fusion Power Associates, a group that promotes fusion energy. It is "almost impossible to imagine initiating a thermonuclear bomb without a nuclear trigger", says Dean. "The energy requirements are just too much; you can't get enough fuel ignited to produce a weapon."

White maintains that the experiments are designed merely to explore basic fusion and plasma science. "The very fact that we're collaborating with the Russians on this indicates that their defence people have totally discounted any military application of this," he says. He acknowledges that the experiments have produced neutrons, as would be expected if fusion had occurred. Further experiments will establish whether the magnetic fields can be used to confine a fusion plasma, he says.

But Theodore Taylor, a retired LANL weapons designer turned anti-nuclear activist, describes Cohen's alert as "appropriate". He adds: "We're both concerned there may be a strong weapons connection" to the LANL research. Taylor says that such fusion experiments will produce interesting physics, but that this is not enough to "balance the danger" of proliferation, particularly given the remote prospects that they could provide a new economic source of energy in the near future.

Christopher Paine, of the Natural Resources Defense Council, says he is less concerned about the immediate threat that the experiments might yield a pure fusion bomb than he is about the risk that other countries will use similar 'peaceful' fusion experiments as a cover for developing nuclear weapons. The information obtained from high explosive-driven implosion experiments can be used in the development of basic fission weapons, he says.

"Our fear is that by not explicitly banning high explosively driven fusion, you create a loophole" in the proposed Comprehensive Test Ban Treaty now being negotiated in Geneva (see *Nature* 381, 267; 1996), says Paine. "By legitimizing this type of research, you are opening up a new avenue of research in states that don't currently perform it."

LANL disputes this assertion, claiming that the experiments are not applicable to fission weapon implosion in that magnetic fields are used to compress directly the deuterium/tritium pellets, whereas fission weapons use a shock wave generated by high explosives to compress the fission core.

But both Cohen and Paine are ambivalent as to whether the experiments should be allowed to continue. Both admit that the cooperation between the United States and Russia on the programme brings political benefits. Russia is also bringing a technology to the table in which it has "a considerable lead", according to LANL.

Recognizing that the programme is unlikely to be stopped, Cohen says that one alternative would be to continue the project, but to give it "the highest degree of classification" in order to prevent the technology falling into the hands of proliferating nations. At present, the research programme is unclassified.

David Kramer

Commercial use of encryption backed

Washington. The government should encourage the unrestricted commercial use of encryption technology within the United States, says a report from the National Research Council (NRC). The report, prepared by an NRC panel chaired by Kenneth Dam, professor of law at the University of Chicago, says that exports of the technology should also be subject to fewer controls. It criticizes the US government's attempts to persuade businesses to use cryptography to which the government has special access. □

MITI turns up heat on research into thermophile genes

Tokyo. Japan's Ministry of International Trade and Industry (MITI) is stepping up its investment in genome-related research by financing the construction of a new research annexe at its Institute for Bioscience and Human Technology in Tsukuba science city, north-east of Tokyo.

The new US\$20-million annexe will analyse proteins translated from the sequenced genes of a deep-sea-vent bacterium. It will also serve as a centre for the culture collection of microorganisms that survive in extreme environments.

MITI already supports a project aimed at sequencing the entire genome of *Pyrococcus shinkai*, a deep-sea bacterium collected from a volcanic vent in the Okinawa trough (see *Nature* 374, 583; 1985). This is one of several recent ventures by MITI into genome research, an area traditionally covered by other ministries.

MITI officials say that the agency is supporting work on thermophiles because it is interesting fundamental research and has the potential to form the basis of new industries. They argue that highly heat-resistant proteins from such organisms may find future applications in devices such as biosensors and biochips.

The new annexe will employ about ten scientists and an equal number of support staff. Michio Oishi, director of the institute, who is also an adviser to the institute at which the sequencing is being carried out, says that work to be done at the annexe will benefit from his institute's existing expertise in protein crystallography.

The new annexe will also function as a centre for the collection, storage and distribution of microorganisms. Oishi says that MITI cannot afford to hire the 50 to 100 staff required to run a large-scale operation like the American Type Culture Collection. He has suggested that the new annexe focuses on what he calls "extremophiles". These are organisms which can survive in extreme temperatures and in very acid or alkali environments, as well as species collected from the inside of nuclear power generators which have adapted to highly radioactive environments.

Such a collection, which would house no more than 200 to 300 species, would complement the other work being carried out on the function of proteins from such organisms. Oishi says that his institute was chosen to house the new culture collection because it already operates a collection of several thousand prokaryotic and eukaryotic cells which have been patented. He adds that he expects the entire genome of *P. shinkai* to be sequenced by the end of next year.

Stephen Barker

Head of climate group rejects claims of political influence

London. The co-chairman of the main scientific panel of the United Nations Intergovernmental Panel on Climate Change (IPCC) this week vigorously defended the organization against suggestions that its scientists have forced through a scientific consensus on global warming to satisfy political demands.

Sir John Houghton, co-chair of IPCC working group 1, which is responsible for reviewing the science of climate change, describes the suggestions as "a mixture of confusion and misinformation". The IPCC's second five-yearly report on climate change, which was published in London yesterday by the Cambridge University Press, is "a unique, very solid and authoritative piece of work, written by the world's best scientists", he says.

After months of detailed negotiations, the report was finalized at a conference last November in Madrid (see *Nature* 378, 524; 1995). But a number of what have come to be known as 'contrarian' groups — representing mainly the fossil-fuel lobby and the energy industry, but also including scientists and economists sceptical of the IPCC — maintain the panel is hostage to politicians and pressure groups convinced, despite the uncertain nature of climate predictions, that global warming exists.

In a statement issued last week, for example, John Emsley, a member of the European Science and Environment Forum (ESEF) — a group of scientists who claim to have been excluded from the IPCC process — accused the panel of censorship. Emsley, a chemist at Imperial College, London, argued that IPCC reports achieved consensus by blocking out alternative views.

Both ESEF and the Global Climate Coalition, a Washington-based fossil fuel pressure group, have repeated allegations that the IPCC 'doctored' its latest report by making last-minute changes to a portion of the approved text, without having the alterations peer-reviewed (see *Nature* 380, 97; 1996). But Houghton says changes were made to background documents only — which IPCC rules allow — and were designed to clarify statements, and not to alter their meaning.

'Contrarian' groups also challenge the IPCC's main conclusion that the "balance of evidence" suggests that human activities are having a "discernible" influence on climate. Emsley says such a conclusion is not sup-

ported by scientific papers in the report. Michael Jefferson, deputy secretary-general of the World Energy Council, an organization that has been closely involved in the IPCC process, says that although changes in climate variability and extremes are beginning to emerge, "global patterns are not yet apparent".

But Houghton, who is to remain in his post for the third five-yearly IPCC assessment, says his critics are wrong. "There are thousands of references to support [the IPCC] conclusions," he says. "It is not an agreement about the detail, because science does not work like that. It is a general consensus about what we know and what we don't know. There is no question of forcing a consensus."

Houghton adds that he is aware of the perception that IPCC scientists are being used for political purposes. "But this is a false perception, and is not just restricted to climate change," he says, arguing that it also applies to areas such as the current debate over bovine spongiform encephalopathy.

"The job of the IPCC is to provide the policy-makers with a scientific message," says Houghton. "We do recognize that the message gets distorted when politics comes into play. That is inevitable. But it is not the scientists' fault."

Houghton says he is aware of political pressures on IPCC scientists and acknowledges that they are sometimes asked by government representatives to make prescriptive statements. "Policy-makers have said to us in the past, 'please tell us what level of CO₂ we should aim for'. They would be delighted if we gave them a numerical answer. But we can't do that."

IPCC reports are written and reviewed by hundreds of authors, including government scientists as well as 'independent' scientists from academic institutions. The ability of government scientists to influence the wording of reports has led some environmentalist groups, such as Greenpeace, to call for measures to minimize political involvement.

Bill Hare, climate policy adviser for Greenpeace, says he is not against government involvement. "But I feel that at present, the balance of influence is too much in their favour."

But Houghton says any move to reduce political involvement in the IPCC would weaken the panel and deprive it of its political clout. "The presence of government scientists is vital to the IPCC," he says. "They own the findings. If governments were not involved, then the documents would be treated like any old scientific report. They would end up on the shelf or in the waste bin."

Ehsan Masood



Houghton: 'solid and authoritative' report.

Cuts spark strike at Australian universities

Sydney. The battle over higher education funding in Australia spilled into the streets last week when academic staff, general staff and students of all 36 public universities staged an unprecedented national strike. About 30,000 protesters took part in demonstrations, and universities were closed by picket lines, prompted by plans for cutting public expenditure being drawn up by the newly-elected conservative Coalition government.

The protest stemmed from the threat of cuts of up to 12 per cent in operating grants, totalling around A\$550 million (US\$433 million) per annum, hinted at by the new federal Minister for Education, Senator Amanda Vanstone, in a meeting with vice-chancellors last month (see *Nature* 381, 265; 1996).

Vanstone has denied that any decisions have been made over the level of cuts. But she is holding firm to an instruction from John Howard, the Prime Minister, that ministers should not reveal where cuts totalling A\$8 billion over the next two years are to be applied until the government's first budget, due in August.

After attacking the vice-chancellors in parliament for taking "an irresponsible attitude" by avoiding their responsibility to contribute to the task of meeting the

government's goal, as well as for "using the meeting as an opportunity to obtain media coverage", she appealed to them to tell her how they would prefer to handle the reduced funding.

But the vice-chancellors have resisted doing what they consider to be the government's work, and have asked Howard for an urgent meeting. Fay Gale, vice-chancellor of the University of Western Australia and president of the Australian Vice-Chancellors' Committee, said in a letter to Howard that the cuts would have a "devastating effect" on universities.

Meanwhile a consortium of science organizations — backed, intriguingly, by the conservative Farmers' Federation — held a joint press conference protesting against rumoured cuts to research funding. Joe Baker, president of the Federation of Australian Scientific and Technological Organisations, who convened the meeting, described cuts to the funding of research and teaching institutions as "like slashing the wrists of the economy".

A leaked government document has suggested that one way of reducing funding of the Commonwealth Scientific and Industrial Research Organization (CSIRO), would be through applying an 'efficiency dividend' of three per cent of its budget,

amounting to A\$12 million. CSIRO has already been subject to a 'dividend' of 1.5 per cent only to its administration, intended to reflect better management and use of technology. Now the threat is to apply an increased dividend across the board, including to research programmes.

Speaking in a television interview, Malcolm McIntosh, the chief executive of CSIRO, said he hoped that the government would not only maintain CSIRO's base funding, but even increase it. "As soon as you do it [a three per cent cut] across the rest of the spectrum, it cuts the science we can do and that is, we believe, contrary to the commitment the government made in its election campaign," he said.

At the same time the business community — ironically the backbone of the conservative government — is itself objecting to reports that a popular scheme of encouraging industrial research and development through 150 per cent tax concessions, introduced by the previous Labor government, is to be watered down. A delegation of research managers emerged from a meeting with John Moore, the cabinet-level Minister for Industry, Science and Tourism, expressing concern that he was not prepared to comment on such reports.

Peter Pockley

Agricultural institutes 'to favour developing countries'

London. The Consultative Group on International Agricultural Research (CGIAR), an international network of agricultural research organizations, backed by the United Nations, last week promised to reorient its research programmes to improve its ability to meet the needs of developing countries.

The move reflects long-standing accusations made against the organization for having put the interests of the industrialized world before those of Third World farmers.

CGIAR was set up in 1971 to conduct research into the cultivation of basic food crops. It is sponsored by the World Bank, the UN Food and Agriculture Organisation and the UN Development Programme. Although 16 of CGIAR's member states are from developing countries, its board is dominated by representatives from its 21 industrialized country members. Indeed, four countries — Australia, Canada, the United Kingdom and the United States — hold half of all key positions in the group.

Environmental groups have frequently claimed that this imbalance biases CGIAR's actions towards the interests of industrialized countries. They have also accused it of having promoted unsustainable agriculture

through, for example, its focus on high-yielding crop varieties and chemical fertilizers. Critics argue that while such an approach has boosted crop yields, it has also

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Weighing up: CGIAR has decided to increase its focus on the needs of developing countries.

reduced the genetic diversity of crop plants and polluted water supplies.

In what seems to be a move aimed at offsetting such criticism, a meeting of the CGIAR in Jakarta, Indonesia last week endorsed a proposal from its Technical Advisory Committee that all research at its 16 centres should be focused on the needs of developing countries, where most of the centres themselves are located.

Ismail Serageldin, the chairman of

CGIAR, says that its research programmes up to the year 2000 "will be skewed in favour of the poor". He adds that the research bodies "would fail in our noble mission if our work was not sharply defined at fighting the misery and hunger of small farmers and landless poor".

Some critics remain unconvinced that this shift in strategy will be translated into action. "We are still very sceptical," says Henk Hobbelenk from Genetic Resource Action International, a non-governmental organization based in Barcelona, Spain.

Furthermore, non-governmental organizations from developing countries remain concerned that the CGIAR's control over germplasm means industrialized countries gain most of the benefits from increased yields and new varieties. "CGIAR has to get its politics of benefit-sharing correct first," says Sunita Narian, deputy director of the Centre for Science and Environment, New Delhi.

But the organization insists that the interests of the developing world have always been its priority. Last week's decision is only intended to give its research a sharper focus, says Heinrich von Loesch, a spokesman for the CGIAR secretariat in Washington.

Anju Sharma

Pollutants 'may harm fetal development'

Washington. Common environmental pollutants may be harming the development of fetuses, infants and young children by interfering with the action of key hormones, a panel of prominent scientists claimed last week. The result, they say, could be reduced levels of intelligence, and learning disorders.

The 18-member panel, which includes scientists in disciplines from toxicology to wildlife endocrinology, drew up its conclusions following a workshop on "environmental endocrine-disrupting chemicals" held last November in Erice, Sicily. They made public their concerns at a press conference in Washington.

The group claims that all pregnant women pass man-made pollutants (called endocrine disruptors) to their fetuses. As

minute concentrations of hormones regulate early development, these could be seriously impeded by even trace amounts of endocrine disruptors that would not harm adults, warns the panel. "There may be no safe level of exposure," it concludes.

As evidence, the scientists cite the structural similarity of certain dioxins and polychlorinated biphenyls (PCBs) to the naturally occurring thyroid hormones that are essential in fetal development. The former have been shown experimentally to compete with thyroid hormone for binding sites on transport proteins, and to alter hormone production and metabolism. Data suggests that they sometimes mimic the hormone's action and sometimes block it. Too much or too little thyroid hormone can

cause problems ranging from motor dysfunction to mental retardation.

"This changes things really by an order of magnitude," says Susan Porterfield, professor of physiology and endocrinology at the Medical College of Georgia, and a member of the panel. "We're talking now about the potential for very small quantities of these compounds having very significant effects," she says, pointing out that the toxicity of other environmental pollutants is proportional to their concentration.

The scientists concede that knowledge of these and other pollutant hormone interactions is still incomplete. But they claim that this example is only the tip of a dangerous iceberg, and that many substances — some unidentified — may also interfere with early development of the brain and nervous system. "It is imperative to monitor levels of contaminants in humans, animals and the environment," the panel concludes.

Endocrine disruptors occur in everything from plasticizers to pesticides, according to the panel. Yet government spending on monitoring them is "trivial". Moreover, the scientists say, the interdisciplinary research needed to increase understanding of the pollutants' effects suffers because of the "narrow expertise" of grant-making scientific panels.

The panel recommends that manufacturers be obliged to name all the chemicals in their products and assure that these pose no developmental health hazards. But this is rejected as unrealistic by Jon Holtzman, a spokesman for the Chemical Manufacturers Association, who argues that it would require companies to divulge proprietary information. US companies already "abide by the most rigorous reporting system in the world", he says.

Meredith Wadman

Health agency avoids Congress's axe

Washington. Attempts by some Congressional Republicans to abolish the US Agency for Health Care Policy and Research (AHCPR), have been blocked in the US Senate, granting a stay of execution that may augur well for the agency's continued survival.

The \$125-million agency carries out research to improve the quality and cost-effectiveness of healthcare. But its involvement in issuing guidelines for clinical practice has made it enemies among some physicians (see *Nature* 377, 379; 1995).

In particular, spine and eye surgeons did not take kindly to the AHCPR's recommendations on when back or cataract surgery is warranted. They lobbied Republicans on Capitol Hill to give them the agency's head on a plate. This request that was duly met in early drafts of both Senate and House budget resolutions for 1997, which would have either eliminated or drastically reduced funding for the agency.

In a bid to forestall such drastic action, the agency told an April meeting of the House appropriations subcommittee that oversees its budget, that it would stop producing clinical guidelines. It promised to restrict its role to acting as a "science partner" with physicians and healthcare planners, conducting background research to help them develop their own guidelines.

The agency's tactics seem to have paid off. Late last month, the Senate's only physician, Bill Frist (Republican, Tennessee) took the floor to proclaim that the agency has seen the light and jettisoned its role of "telling [physicians] how to practise medicine". He added: "We should applaud this type of initiative and responsiveness, not cripple it," he said, a reference to the budget resolution which cut nearly two-thirds of the agency's budget.

Pete Domenici (Republican, New Mexico),

chairman of the budget committee, promptly restored full funding for the AHCPR. The parallel House resolution, also revised from early drafts, recommends killing only the agency's clinical guidelines work, but makes no specific funding cuts. All told, the agency has survived a "close call", admits Clifton Gaus, its administrator.

But the battle is not yet over. The budget resolutions serve only as guidance for 13 appropriating subcommittees which are now drafting 1997 spending bills. Sam Johnson (Republican, Texas), the chief critic of the agency in the House, recently told the relevant subcommittee that the agency is a "boondoggle", and promised that he would "still advocate eliminating funding for the agency". But, Gaus says that he remains "optimistic" about the future of the agency.

M. W.

India's missile centre shaken by blast

New Delhi. The Indian government has ordered a high-level investigation into a huge explosion and fire which occurred last week at its top missile research centre, outside Hyderabad. Sabotage is not being ruled out, given the recent terrorist bombings of a shopping centre in Delhi and a tourist bus in Rajasthan.

The Defense Research Development Organization immediately denied news reports that up to 20 people had been injured in the blast, which was heard ten km away. A. P. J. Abdul Kalam, the rocket designer who heads the organization, also claimed that the incident would not have any "adverse effect" on India's missile programme. But Kalam said that safety measures at the complex would be tightened, whatever the outcome of

the enquiry into the explosion.

The damage, described as "heavy", was confined to the laboratory for developing chemical additives for propellants in the High Energy Materials building where the explosion occurred, according to Kalam. "The main laboratory a few metres away, where missiles are assembled, escaped unscathed," he said.

The research centre is central to India's plans to develop medium and long-range missiles, including the surface-to-surface Prithvi missile. The programme is controversial, anti-nuclear organizations arguing that the missiles could be fitted with the nuclear warheads that India is believed to have developed following its underground nuclear test in 1974.

K. S. Jayaraman

Paper from jailed professor stirs debate over publication

Montreal. A materials research journal has refused to publish a paper written by an engineer from the prison cell in which he is serving a life sentence — with no eligibility for parole for 25 years — for murdering four of his colleagues in 1992, stirring a controversy over whether the decision is an infringement of academic freedom.

The *International Journal of Solids and Structures* published an earlier article from Valery Fabrikant in January. But its editor, Charles Steele, a professor of applied mechanics at Stanford University, has now bowed to pressure from university staff and students, as well as the family of one of Fabrikant's victims, not to publish any more of his papers.

After Steele published Fabrikant's earlier paper, "A Solution to a Complex Problem on Cracks", he received a letter of protest from Frederick Lowy, rector of Montreal's Concordia University, and Donat Taddeo, its dean of engineering. Both argued that Fabrikant had forfeited his liberty, and with it the right to publish his academic views.

Steele admits that he has been "in a quandary" as to what action to take. Many of his colleagues advised him that it was correct to publish the second article, and he agrees that the article was "pretty sound". He concedes that his final decision was somewhat "arbitrary", and that he was swayed by objections from the family of one of Fabrikant's victims and from graduate students in his own faculty.

But Steele's action has come under fire from Margaret Somerville, director of the McGill University Centre for Medicine, Ethics and Law. "The best test of a society is how it treats its most despised members," she says.

Somerville argues that society should not allow an individual's crimes to influence judgements on the science produced by such individuals unless the two are directly related — as was the case of Nazi experiments on humans. "It's tremendously important that we don't start putting irrelevant characteristics in judging what we will and will not publish scientifically. The big question is: is what he did relevant to the scientific integrity of his work?"

Lowy and Taddeo, in their letter to Steele, also claimed that Fabrikant, whose ability to influence the opinions of others had been apparent both at the university and during his trial, might be trying to exploit the recognition conferred by scientific publication to improve his prospects for parole. But Somerville argues that the significance of published papers in terms of parole prospects is a matter for a parole board, not a journal editor.

Bernard Dickens, a professor of law at the University of Toronto who is attached to its Centre for Bioethics, agrees. "If the content of the article is sound, it should be published," he says. "It would be inconsistent with the goals of a university to attempt to suppress knowledge." **David Spurgeon**

Russian rescue effort nears polar scientists but faces icy barrier

St Petersburg. One hundred and twenty-two Russian polar explorers, who have been stranded in Antarctica for over a year and are down to their last stocks of food, could — weather and mechanical reliability permitting — soon be heading for home. The explorers' plight is the result of the Russian government's inability to pay for repairs to the ice-breaker that broke down three times on its way to rescue them.

The 'changing of the guard' in the Novolazarevskaya research station usually takes place three times each year — in December, February and once during the summer months. The Russian government maintains a fleet of ships that are used to deliver equipment, food and fuel, as well as a change of personnel. But for the past year, the government has only been able to pay for one ship, and has been unable to set aside funds to pay for repairs.

When the ship, the *Academician Fedorov*, broke down for the first time, it was taken to a dock in Germany for repairs, and was then duly impounded when the Russian Arctic and Antarctic Institute (AAI) was unable to foot the repair bill. The ship was eventually released in January, after the dockyard had secured half the repair costs, plus a promise that the remainder would eventually be paid.

The second breakdown occurred as soon as the ship had crossed the Equator. *Academician Fedorov* remained in Cape Town for three weeks while its engine underwent further repairs. After a third breakdown, the ship was directed to Buenos Aires, where repairs took a further month. The ship has now left the Argentine capital, and is on its way to the Novolazarevskaya polar station.

Even now, the success of its voyage is not guaranteed, as most of the favourable weather has been missed and the ship will have to cut through heavy Antarctic ice in the polar night. If conditions prove too difficult, the explorers will be evacuated by helicopter.

The explorers, meanwhile, are said to be bearing up well, but they have stopped all scientific activity. Food stocks are low, and fuel is expected to last two to three weeks, according to Arkadyi Soshnikov, director of the AAI. But, he says, the researchers are not panicking. "They took part in a recent poll to elect a governor for St Petersburg, and are making preparations for voting in the Russian presidential elections on 16 June."

Carl Levitin

Universities charge for Internet access

San Francisco. Californian colleges and universities have begun to charge faculty members and students for external access to the Internet. They say the move has been made necessary by unmanageable growth in the volume of on-line traffic, and the resulting increased costs (see *Nature* **380**, 377; 1996).

Almost two-thirds of traffic on the University of California's campus networks is now caused by students and staff dialling in from off-campus locations. As a result, links have become congested and slow, while lines are often engaged. Much of the increase in traffic is due to a shift in Internet use from educational and research purposes to entertainment and other activities, in particular using the World Wide Web.

For students and faculty members at the University of California at Berkeley, for example, such dial-in connections are no longer free, with users now being

required to pay \$10 a month to a commercial Internet service provider (the service will remain free on campus). Other UC campuses and California state universities are beginning to follow suit.

Stuart Lynn, associate vice-president for information resources and communications for the University of California's Office of the President, says the increasing demand had made it impossible for the university to continue providing the service free. Other universities across the United States, including Cornell University, have taken similar action, he says.

The university says it has received few complaints about its new policy. But it admits some academics have expressed concern about the weakening of the tradition of free access to information and educational tools, and have been warning against the gradual invasion of commercialism into collegial values and the sharing of information. **Sally Lehrman**

California looks to science partnerships ...

Sacramento, California. A formidably powerful science and technology base is helping to lift California out of its longest recession since World War Two. But new types of partnerships between science, industry and local government are needed to sustain the state's recovery, according to participants attending the California Coalition for Science and Technology Summit, held in Sacramento last week.

The meeting, which was organized by the University of California at Davis and sponsored by Californian corporations and local and federal agencies, failed to reach a clear consensus on what California should do to reinforce its science and technology base. But it did agree that various parts of the state needed to work harder to emulate the world-beating performance of its strongest regions — Silicon Valley and San Diego.

The meeting's recommendations were:

- Less resistance from universities to direct contact between faculty and industry. "The universities have not done a very good job" in technology transfer, admits Richard Atkinson, president of the University of California (UC) system. State laws inhibiting the commercial activities of UC faculty are among the most restrictive in the United States, other participants said.

- Better technical education in high schools. Delegates raised this issue more often than any other — but when Dan Goldin the administrator of NASA, asked how many people in the hall spent time helping schools, scarcely a hand went up.

- Stronger regional, non-profit groups to foster ties between universities, government laboratories, industry, local government and venture capitalists.

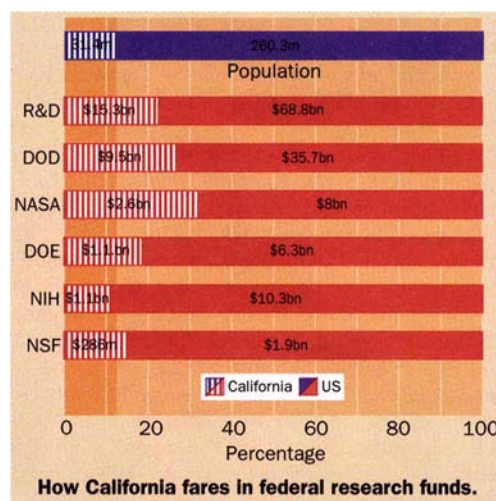
- Cheaper means of accessing science inside government research laboratories. The giant Lawrence Livermore National Laboratory (LLNL) near Oakland was praised for its expertise, but criticized as "bureaucratic" and expensive to work with.

- Network, network and network. AnnaLee Saxenian, an economist at the University of California at Berkeley, pointed out that an open and informal networking system is the basis of "Silicon Valley's unique dynamism". Traditional science-industry links, such as those between the California Institute of Technology and the aerospace and electronics industries in the Los Angeles area, were less suited to the needs of today's economy, she said.

Environmental technologies — which only two years ago were being touted as a saviour for Lawrence Livermore — were scarcely mentioned at the meeting. The impact of the loss of defence-related jobs, which dominated California's economy for fifty years, seems to have been overcome as the state races into a boom based around silicon, biotechnology and the film industry.

"A very, very dramatic recovery" delivered a net gain of 300,000 jobs to California last year, says Tapan Munroe, chief economist at Pacific Gas and Electric. "The renaissance of California is being driven by science," was the economist's bullish verdict.

Although the state houses only 12 per cent of the US population, it attracts one-quarter of federal research and development spending (see figure). It also hosts the world's three largest biotechnology companies — Chiron, Amgen, and Genentech —



and biotechnology provides 165,000 jobs, at an average annual wage of \$50,000.

In the new multimedia industry, California has more young firms than the rest of the United States combined. As an omen for the future, it is studded with "little giants" such as Netscape, the Internet services company, which already has a stock market value of \$4 billion.

Against this optimistic backdrop, Ernie Moniz, associate director for basic science at the White House's Office of Science and Technology Policy (OSTP), warned that California could be hard hit by cuts in the research activities of the federal government. Congress's plans for this year could cut \$100 million in California alone, he said.

And some support was voiced for a mechanism to prioritize science and technology spending in the state. Dan Thompson, of the manufacturing technology programme at LLNL, points out that California has "no broadly based structure to make resource decisions" for research and development. "There are even some who say it doesn't need it," he says.

One supporter of this point of view is Charlene Simmons, who works for the research service of the state government in Sacramento. "Maybe we're better off without a state science and technology programme," says Simmons, adding that the state civil service is "too slow" to administer such a programme efficiently.

But informal coordination, probably orchestrated through local groups, will definitely increase as people learn from the trauma of the recent slump. "The stress has opened people's eyes to what has to be done if they're not going to be stressed again," says M.R.C. Greenwood, chancellor-elect of the University of California at Santa Cruz, who came up with the idea of the summit during her period at the OSTP. "This felt like the right meeting," she says. "Its success will encourage more participation in partnerships".

Colin Macilwain

...and supports medical research

Sacramento. Californians overwhelmingly approve of federal support for biomedical research, and many would even pay higher taxes to fund more of it, according to a public opinion poll. But only one in twenty-five of those interviewed identified correctly the National Institutes of Health (NIH) — which has an annual budget of \$12 billion — as the government organization that pays for the research.

The obscurity of NIH contrasts sharply with the celebrity of NASA, which 53 per cent of those polled identified as the space agency, and the Food and Drug Administration (FDA), which 67 per cent identified as the body approving drugs and medical devices.

"Historically, NIH has been comfortable with its role in the background," explains Ed Penhoet, president of Chiron, the biotechnology company.

Asked if they approved of federal funding for medical research in universities, 86 per cent of those polled said yes, and 51 per cent thought it should be increased. Seventy per cent considered the federal dollars already spent good value, a slightly higher proportion than those who felt the same about spending on defence (64 per cent), social security (62 per cent) and public education (55 per cent).

Over half said that in principle they would even be willing to pay an extra dollar a week in taxes to support medical research. Three quarters of those questioned supported the use of animals for medical research, while one-fifth opposed it. The poll was commissioned by Research!California, a branch of Research!America, a group which aims to foster public support for science.

C. M.

Japanese minister takes pay cut to atone for HIV contamination

Tokyo. In an unusual attempt to atone for his ministry's failure to prevent the spread of HIV among Japanese haemophiliacs in the early to mid 1980s, Japan's Health and Welfare minister, Naoto Kan, last week announced pay cuts for himself and other top ministry officials. Kan and vice-minister, Hiroshi Tada, will take a 20 per cent cut in pay over the next two months. In addition, 13 other senior officials have received severe verbal reprimands for their handling of the disaster in which about 2,000 of Japan's haemophiliacs were infected with HIV from non heat-treated blood coagulants.

Kan, who only became minister in January this year, bears no responsibility for the disaster, and has spearheaded moves to force exposure of the ministry's culpable actions over the past 13 years. But it is in line with Japanese tradition that Kan, no matter how blameless, should take symbolic responsibility. Representatives of haemophiliacs, however, expressed dissatisfaction with the ministry's action, pointing out that a 20 per cent pay cut for two months is hardly adequate atonement for the infection of 2,000 people, of whom about 400 have died. □

Merck signs diabetes pact

New Jersey. The pharmaceutical company, Merck & Co. Inc., has signed a collaborative research and licensing agreement with the Wellcome Trust and the University of Oxford for work on the discovery of human genes associated with the development of Type I — insulin-dependent — diabetes. The terms of the agreement have not been disclosed, but it represents Merck's first external collaboration in the area of genomics with the intention of identifying specific genes related to a disease.

The work will be carried out at the Wellcome Trust Centre for Human Genetics, which was established by the University of Oxford and the Wellcome Trust. The centre has pioneered the identification of genetic markers that show a relationship between Type I diabetes and a physical location on a chromosome. □

German reactor gets go-ahead

Munich. Germany's next-generation fusion reactor, the Wendelstein-7X, designed by the Max Planck Institute for Plasma Physics in Garching, last week received the official go-ahead from the federal and *Länder* governments, which are to share its DM500-million (US\$333-million) investment costs with the European Union. The final signing of agreements after months of uncertainty (see *Nature* 378; 652; 1995) means that Wendelstein-7X, to be sited at Greifswald in East Germany, should begin to operate by 2005. □

Habitat II urged to heed science

Istanbul. Scientists representing 72 of the world's academies of science have urged world leaders to give science and technology a higher priority in solving urban problems and in reconciling economic advancement with environmentally sustainable development.

In a statement issued to coincide with the United Nations Conference on Human Settlements, Habitat II, the representatives expressed concern that science and technology issues were being ignored in a conference document that is likely to guide future policy on urban development and the problems faced by 'megacities'.

Science and technology have a crucial role in ensuring the long term sustainability of cities, said F. Sherwood Rowland, foreign secretary to the US Academy of Sciences, a Nobel laureate and co-chair of the inter-academy panel that co-sponsored the Istanbul meeting. "But it can't be done unless you have the education and training in place to ensure that scientific capacity continues to grow. And that concept, as well as others regarding science and technology, is missing from the document." □

Psychologist must 'modify style'

London. Chris Brand, a psychology lecturer at the University of Edinburgh and the author of a controversial text on race and intelligence, has been asked by the university to "modify his style of teaching" following complaints from students (see *Nature* 381, 105; 1996). An inquiry into Brand's teaching conduct concluded that his lectures needed to be more balanced and his relationship with students one of "mutual respect".

A statement from the university said that the psychology department "will be asked to consider redistributing the teaching load in certain areas [such as intelligence and personality] which to date have been highly concentrated upon Mr Brand". The inquiry noted, however, that Brand was fair and impartial when he was assessing students' work. □

Woman heads Indian biotech unit

New Delhi. Manju Sharma, 53, has become India's first woman scientist to be given the rank of secretary in the Ministry of Science and Technology. She has been appointed secretary to the Department of Biotechnology (DBT), a key department in the science ministry, succeeding Chittaranjan Bhatia, who retired five months ago. Until her selection last week, Sharma was an adviser in the DBT. A botanist by training, she is also president of the National Academy of Sciences in Allahabad. Sharma said that DBT will give priority to applications of biotechnology to agriculture and to developing DNA technologies for prenatal diagnosis of genetic disorders. □

China and Germany plan links

Bonn. China and Germany last week signed an agreement that provides for the extension of cooperation in scientific research and education. The Chinese minister for State Education, Zhu Kaixuan, and his German counterpart, Jürgen Rüttgers, agreed that moves to enhance links between the two countries should be completed within three years, after which further links to the European scientific research community will be considered. The Chinese Educational and Scientific Research Network and the German Research Network Association have been charged with working out the details of the agreement. □

Radar array will enlighten aurora



London. Greater understanding of the Northern Lights, or Aurora Borealis, which are caused by storms on the surface of the Sun, will soon be possible thanks to the completion of a new radar interferometer consisting of antennae in Finland and Iceland. The Collaborative UK Twin Location Auroral Sounding System — or CUTLASS — has 16 antennae in its main array, capable of viewing 3 million square-kilometres, an area equivalent to the size of western Europe. It was designed and constructed at Leicester University, with financial support from the UK Particle Physics and Astronomy Research Council, as well as from the governments of Finland and Sweden. □

Will space-based astronomy give value for money?

Astronomers and space agencies are planning missions that will not be launched for twenty years. Before that, spacecraft will have opened up more regions of the electromagnetic spectrum to space-based observation. Astronomers' practices will have to adapt to tight budgets if they are to get the best return on investment.

INTRODUCING the National Research Council's "decadal" list of priorities for US astronomy and astrophysics in 1991, John Bahcall of the Institute for Advanced Studies added a caveat: "Everyone will be deliriously happy and astonished if we get everything we wanted." Five years later, not only has the plan not unravelled completely, like so many other wish-lists in science, but its agenda for space-based astronomy remains remarkably on track.

The Bahcall panel's highest-priority large space project, the Space Infrared Telescope Facility (SIRTF), will be leaner and later than originally planned, but a launch in 2002 now seems assured. Most of the smaller space-based projects recommended by the panel — including the Stratospheric Observatory for Far-Infrared Astronomy (SOFIA) and new Explorer missions for astronomy — are solidly in the pipeline.

Europe, too, is holding to an ambitious agenda for space astronomy in the next decade, embodied in its merged Horizon 2000 and Horizon 2000+ programmes. Despite recent cuts to its science budget (see *Nature*, 379, 476; 1996), the European Space Agency (ESA) has approved two large 'cornerstone' missions for investigating the X-ray and infrared/submillimetre areas of the spectrum, as well as medium-sized satellites to explore the infrared and gamma-ray regions and the primordial radiation left over from the Big Bang.

To these projects one can add a series of Japanese X-ray, infrared and radioastronomy satellites, a trio of Russian space observatories, and a handful of smaller projects from nations as diverse as Sweden and Argentina. Altogether, more than 30 space astronomy missions, costing well over \$1 billion a year worldwide (or ten times the US National Science Foundation's annual budget for astronomy), are planned for the next decade (see chart, page 462). And that does not include the many projects planned for planetary exploration and solar-terrestrial studies, which fall outside the scope of this article. Not bad for an era of declining budgets.

It seems fair, then, to look critically at the return on that investment. Are any of the proposed projects — particularly those of competing nations — wasteful or duplicative? Are the data well used? And what great science are we expecting?

If there is waste, it is certainly on a smaller scale than in the past. The days of billion-dollar satellites are over. The US National Aeronautics and Space Adminis-

tration's four "Great Observatories" — a concept rooted largely in politics as a way to seek unified funding in the 1980s — almost foundered halfway through, when, in the words of Claude Canizares of the Massachusetts Institute of Technology, an investigator for the Advanced X-Ray Astrophysics Facility (AXAF), it "stopped being politically correct" to propose large, expensive space missions. AXAF and SIRTF were first threatened with cancellation, then cut back — money for SIRTF fell from \$2 billion to \$450 million. Smaller missions also felt the pinch — the \$250-million Far Ultraviolet Spectroscopy Explorer had to cut its cost by more than half and shorten its development schedule by two years to stay alive. Now it will launch in 1998, the same year as AXAF.

Even scaled back, SIRTF is probably the most eagerly awaited astronomy satellite in the queue. There is scarcely an astrophysical question it will not address, from the search for proto-planets — objects that are growing into planets by accreting matter — to the study of the very earliest galaxies, whose

visible light, as a result of the Universe's expansion, is shifted to infrared wavelengths that are beyond the range of the Hubble Space Telescope.

So rich is the infrared region scientifically — and so dramatic the recent improvements in infrared detectors — that astronomers are now calling for the 'Next Generation' Space Telescope to be designed so as to target the infrared, rather than visible/ultraviolet, spectrum. The Hubble Space Telescope will extend its capability into the infrared next year, when astronauts install the new Near-Infrared Camera and Multi-object Spectrometer. The European-built Infrared Space Observatory (ISO) is already returning its first results, with big dividends expected.

NASA/SPL

A host of other approved projects will join the infrared astronomy boom, including SOFIA, the US/German successor to the Kuiper Airborne Observatory, and a small NASA project called WIRE (Wide-Field Infrared Explorer), which will scour a small patch of sky for galaxies with high rates of star formation. Japan's Institute of Space and Astronautical Science is placing high priority on the proposed IRIS spacecraft, which will test mechanical cooling techniques. The latter development could liberate space-based infrared astronomy from the reliance on ultra-cold liquid coolants that limits an orbiting telescope's useful lifetime.

The 1990s will also see one part of the electromagnetic spectrum — the submillimetre regime — opened to space-based observation. The most ambitious of these projects will be ESA's Far Infrared and Submillimetre Space Telescope (FIRST), scheduled for launch in 2005. FIRST promises a rich scientific yield — revealing the ►

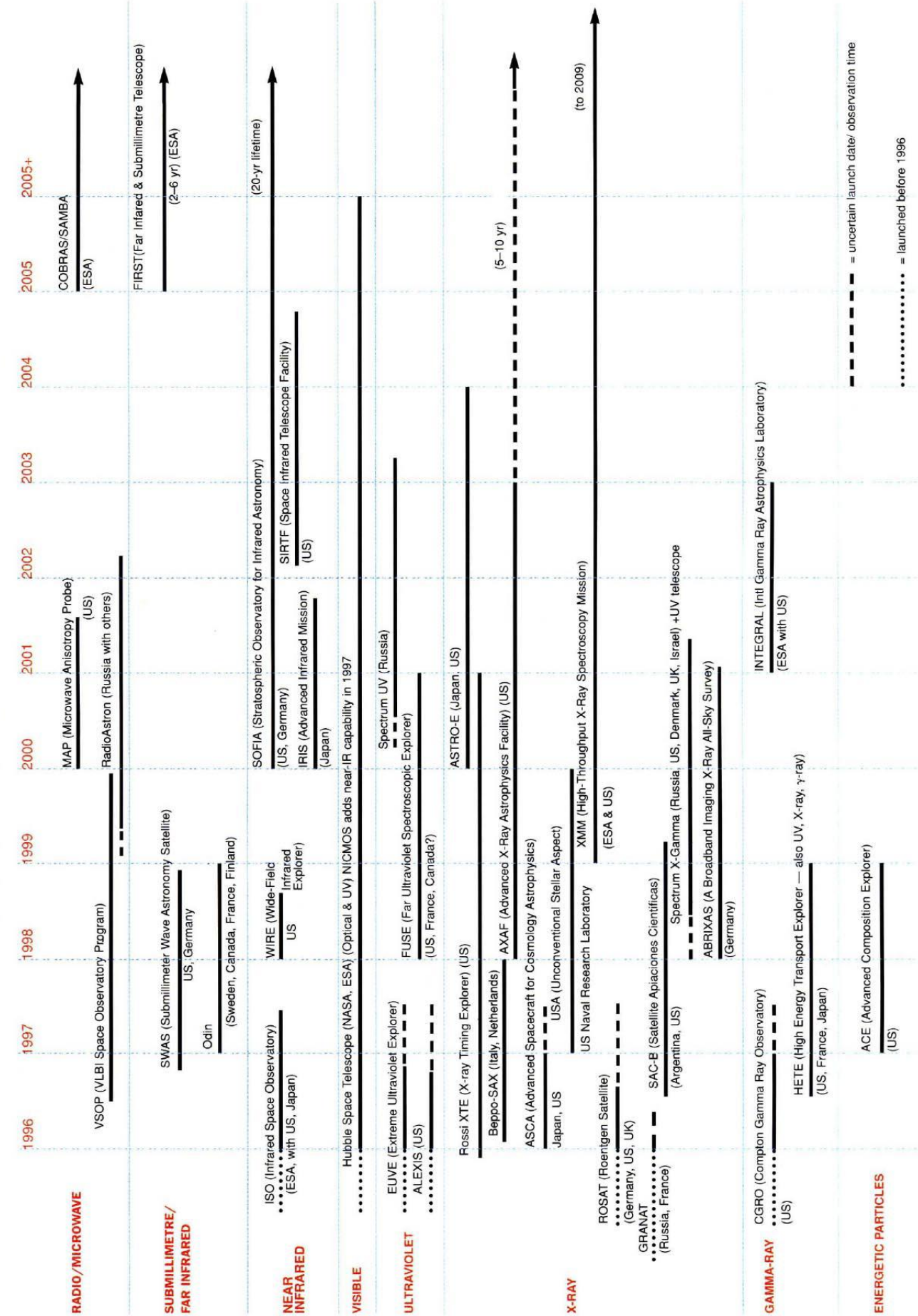
Repairing the Hubble Space Telescope — great but costly science.

On other pages

- Space observatory missions 462
- A new era for radioastronomy 463
- What is next on the horizon? 465

For reprints of this briefing, contact Ross Sturley, *Nature*, 4 Crinan Street, London N1 9XW, UK.

NATIONAL AND MULTINATIONAL ASTRONOMY SPACE MISSIONS PLANNED FOR THE NEXT DECADE



► details of star formation and interstellar chemistry — to justify its price tag of \$800 million. But it also illustrates the opportunities and difficulties of one means of enhancing value for money, that of international collaboration.

ESA's normal policy for expensive cornerstone missions is to rule out significant non-ESA participation. This is largely a matter of regional pride — a principle of space-science autonomy — but is also a way of protecting itself against a partner dropping out during the project. In 1993, however, ESA realized it would need help to pay for the FIRST mission, and so invited NASA to come in with a contribution of \$150 million.

NASA has asked a team headed by Thomas Phillips of the California Institute of Technology to assess how the United States might participate. One idea would be for NASA to provide a boost to a much higher orbit — at a 'Lagrange point' where the balanced gravitational influences of the Earth and Sun provide a stable orbit — far from the Earth's reflected radiation, and boosting the sensitivity. Another would be for the United States to contribute technology that would allow an increase in telescope diameter.

Infrared Space Observatory: technicians work on the rear of the primary mirror, which is cooled by liquid helium.

Radioastronomy floats into space on VSOP

THE launch later this year of a small Japanese satellite with the unassuming name of MUSES-B marks an important advance in radioastronomy, and a new level of sophistication in the international coordination of space projects.

The satellite, to be launched in September by a new solid fuel rocket developed by Japan's Institute of Space and Astronautical Science, will be the world's first orbiting radio telescope for Very Long Baseline Interferometry (VLBI), in which information from widely separated telescopes can be combined to give much increased resolution. Circling between 1,000 and 20,000 kilometres above the Earth, the 800-kilogram satellite with an 8-metre deployable antenna will be linked to a network of dozens of ground-based radio telescopes scattered around the globe to create by far the longest baselines ever used for VLBI — up to 30,000 kilometres, or three times the baselines achievable on Earth.

Years of preparation have gone into Japan's VLBI Space Observatory Programme (VSOP) project, which involves scientists and observatories in the United States, Europe, Australia and many other countries. A similar project called RadioAstron is also being planned by Russia in collaboration with many of the same scientists and observatories. Launch is tentatively marked for the end of this decade.

The long baselines that can be obtained with VSOP will give a much higher angular resolution than Earth-bound VLBI networks — 0.1 milliarc seconds at VSOP's highest observation frequency of 22 GHz. This will allow, for example, detailed imaging of the centres of active galaxies, as well as studies of spot sizes in maser sources, and high-resolution imaging of stars that are unusually strong emitters of radio waves. By teaming up with the Australia Telescope in the Southern Hemisphere, the project offers the first all-sky mapping

instrument for radioastronomy.

VSOP was once just a gleam in the eye of a few Japanese scientists, in particular Masaki Morimoto of the Nobeyama Radio Observatory. But now VSOP and RadioAstron constitute a "great big amorphous international project," according to Jim Ulvestad of NASA's Jet Propulsion Laboratory in Pasadena, California, which is providing crucial support for both projects.

Ulvestad, who for two years co-chaired the International Mission Operations Group that helped to coordinate the two projects, says that while the technical aspects of space VLBI are difficult, "the coordination and negotiation among all the international partners is by far the most difficult thing to do." The Internet and e-mail are essential tools, but there is "no completely satisfactory substitute for face-to-face meetings", says Ulvestad.

Overseeing and managing this coordination are two sets of international committees, with overlapping members drawn from all key participating organizations. The VSOP Science Operations Group and the RadioAstron equivalent are responsible for mission planning and operations. They will also ensure there is no conflict in the use of ground facilities if the two satellites are in operation at the same time. These groups are overseen in turn by two international science councils which set major policy and scientific goals. Independent science review committees have been set up to screen proposals, while Japanese and Russian teams are responsible for day-to-day operation and control of the satellites.

Since space VLBI depends equally on space- and ground-based observations, VSOP and RadioAstron have stimulated a flurry of activity at radio observatories around the world, upgrading facilities to make them more compatible with the two satellites. The European VLBI network, a consortium of observatories operating radio telescopes in Britain, Sweden, Ger-

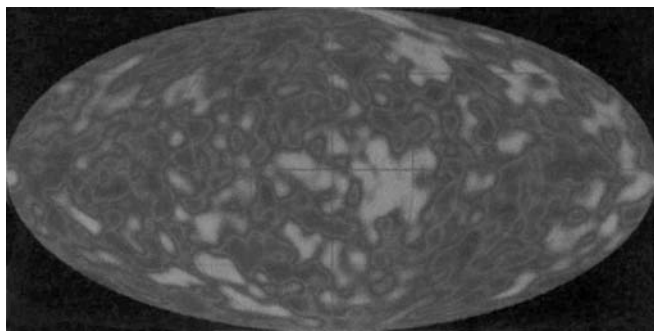
many, the Netherlands, Italy, Spain, France, China, Finland, Poland and the Ukraine, is upgrading all telescopes to conform with the wide-band capability required for VSOP. This involves better tape recorders, new electronics, new filters and new control software.

But money problems hamper RadioAstron's prospects of launch. There have already been delays in payments for the project, and two other large Russian missions, Mars 96 and Spectrum X-Gamma, are in the queue ahead of it. As a result, there can be "very little confidence... in any particular launch date", says Leonid Gurvits of the Joint Institute for VLBI in Europe, who is also affiliated with the Astro Space Center in Moscow, the organization leading RadioAstron. Still, given the high level of readiness of most of the on-board science payload, all the necessary ground infrastructure, and the very extensive technical expertise in the Russian space complex, Gurvits is confident that "with appropriate funds" RadioAstron can be launched by the end of this decade.

RadioAstron will be much bigger than VSOP, weighing 5,000 kilograms. It will also have a larger (10-metre) antenna, and its orbit will extend farther — 77,000 kilometres, giving longer baselines and up to three times the angular resolution of VSOP. Hisashi Hirabayashi, project leader at the Institute of Space and Astronautical Science, claims, however, that VSOP should give better-quality images.

Even before VSOP's launch, a more ambitious follow-up mission is being planned. Both VSOP and RadioAstron are less sensitive than typical ground-based radio telescopes, partly because of the small size of their antennas. That problem may soon be solved. Last month, the US space shuttle deployed and tested a 14-metre inflatable antenna structure whose successors might one day extend the frequency range of the first generation of radio telescopes in space. □

► The benefits of cost-sharing on the FIRST mission are obvious. But major US participation would also be a gamble for ESA because of the unpredictability of NASA appropriations. Phillips admits to the difficulty, saying, "it is a question of trust."



COBE's microwave background — wonderful but blurred.

And ESA's principle of autonomy holds true, in that the mission can exist without the US involvement if necessary.

In general, according to astronomers, the proposed international projects do not duplicate one another — too much money is at stake simply to repeat someone else's experiment. The US AXAF and European High-Throughput X-Ray Spectroscopy Mission (XMM) spacecraft, for example, will launch within a year of each other, and will cover the same region of the spectrum. But the US spacecraft will have the better resolution and its European counterpart the higher sensitivity — competition, but also complementarity. Spacecraft effort is booming in the X-ray and ultraviolet region, because these wavelengths are inaccessible from the ground.

Microwave mappers

Considering the expense of reaching orbit, one might ask if we can afford any two spacecraft filling the same niche. Recently, a pair of missions was announced which will build on the famous success of the 1992 Cosmic Background Explorer experiment, which mapped the faint glow of microwave radiation left over from the Big Bang. NASA's \$100-million Microwave Anisotropy Probe (MAP) and ESA's \$450-million COBRAS/SAMBA mission will be competing to see who can answer one of modern astronomy's great mysteries — when did structure first appear in the Universe? In case anyone misses the competitive spirit, the COBRAS/SAMBA page on the World Wide Web advertises it as "the definitive cosmic microwave background imaging experiment".

Not surprisingly, cosmologists are happy to have two microwave background missions instead of one. Dick Bond of the Canadian Institute for Theoretical Astrophysics says that both MAP and COBRAS/SAMBA, with their different frequency ranges, are necessary, together with continuing ground-based observations, to crack a difficult problem. If COBRAS/SAMPA were already up,

MAP would be hard to justify, since it is less efficient at probing the microwave spectrum. But MAP will fly in the year 2000, five years ahead of its larger European counterpart.

COBRAS/SAMBA is expected to go further than MAP in yielding clear-cut exclusions of cosmological models, and key quantitative information on the nature of cosmic irregularities occurring when the cosmic background radiation was being generated — less than a million years after the Big Bang.

Where pride and politics play less of a role, collaborations can be found in plenty.

Most of the planned space astronomy missions, in fact, have some international cooperation, whether in the form of a single guest investigator, a full partnership or some kind of trade. Currently lacking a near-term infrared mission of its own, NASA helps with ground tracking for Europe's ISO observatory in exchange for observing time. Japan's ASTRO-E X-ray satellite provides a refuge for an important instrument that was jettisoned when AXAF was reduced in scope. The Russians, who plan a scaled-down Great Observatory programme with the Spectrum X-Gamma, RadioAstron and Spectrum UV projects, have struck a deal with NASA to help format and archive the Spectrum X-Gamma project data, which will also be handled in Europe.

Most notable of all is the teaming of radioastronomers around the world to extend interferometry to space (see box, page 463).

Having invested billions in these projects, what kind of return might we expect? Some astronomers still see ground-based and space-based astronomy as being in opposition, and grumble that one \$5-billion Hubble Space Telescope could have paid for 50 ground-based Keck telescopes. It is accepted that mission costs must be reduced, however, and the distinction is becoming less important, while opportunities for collaboration abound. NASA has put money into the Keck II telescope to support the agency's search for other planetary systems. Managers of Germany's ROSAT X-ray satellite have an agreement giving ROSAT users access to telescopes owned by the

European Southern Observatory for follow-up observations of interesting X-ray sources.

Most European space astronomers would like to see such relationships expanded. Giovanni Bignami, an investigator for the XMM X-ray mission, has campaigned for an agreement with ESO's planned Very Large Telescope in Chile, and informal discussions between the two organizations have started. "It's the obvious way to go for the greater benefit of both communities — which are in fact only one," he says.

Spreading the bytes

Another way of examining the value of space-based astronomy is to gauge how effectively it shares the wealth of data returned by satellites. Here, too, there are signs of change. NASA missions like the Rossi X-ray Timing Explorer are shortening or ending altogether the 'proprietary' periods during which project scientists can hold on to data for their own use. ESA, while continuing with one-year periods of data protection for those allocated telescope time, is, for the first two years after launch, making 40 per cent of the XMM observatory's viewing time 'guaranteed' for scientists involved in developing the mission. Apart from the additional motivation this gives project astronomers, it ensures that those who best understand the instruments get a chance to use them. But astronomers acknowledge a need to ensure that those using guaranteed time carry out science comparable in quality to that conducted, after peer-reviewed competition, in the other 60 per cent of observing time. ►

The Keck telescope — a ground-based collaborator.

What will be the Next Big Thing?

It does not take much to stir Dan Goldin to grand visions, and at last winter's American Astronomical Society meeting in San Antonio, Texas, the NASA Administrator was in rare form. He had just learned that a panel headed by Alan Dressler of the Carnegie Observatories had proposed an orbiting four-metre telescope as the logical successor to the Hubble Space Telescope. Wonderful, Goldin told the assembled astronomers, except that he thought Dressler was thinking too small.

Goldin's audacity drew a laugh from the crowd. Within a few months, however, NASA was talking seriously about placing not a four-metre, but an eight-metre, observatory in space by 2005. The result would be a space-based instrument with a mirror the size of the largest ground-based optical telescopes, such as the two that will form the Gemini telescope, only a few years after Gemini itself opens for business on the ground.

The Dressler panel's "HST and Beyond" report, which involved astronomers on both sides of the Atlantic, makes several recommendations (see Commentary, p.470), including operating the current Hubble Space Telescope in a reduced mode beyond its planned termination date in 2005, and building an optical interferometer in space. The plan's centrepiece, though, is the Next Generation Space Telescope (NGST), which would be designed to target infrared observations rather than the visible/ultraviolet regime explored by the current Hubble telescope. Such an instrument would be especially powerful for studying the origin of galaxies.

Even before the report had been finalized, NASA was acting on it. A study group at NASA's Goddard Space Flight Center and the Space Telescope Science Institute has been working for weeks on the idea of a next-generation telescope, and recently signed two industry teams to conduct their own feasibility studies. Already, says John Mather of Goddard, the NGST study scientist, the group knows enough to have "great confidence we can build this thing."

The requirements are stringent: only \$500 million for an eight-metre instrument orbiting at one of the Earth's Lagrange points, where it would be naturally cooled to around 30 degrees Kelvin — low enough to cover the 0.5- to 5-micrometre visible/infrared range. Advanced cooling technologies could push the range to 40 micrometres. Launch costs, and hence weight, must be minimized, so the telescope would have a super-thin (a few millimetres thick) deployable mirror. These

are challenging technologies, say the NGST team, but all within reach.

There are sceptics, of course. And, says Mather, "there should be. We haven't built anything yet." Marc Davis, co-chair of the National Research Council Committee on Astronomy and Astrophysics, counts himself among those who would love to see an NGST, but says, "I'd be flabbergasted if it turns out to be cheap."

Mather says NASA is counting on more than technological magic to ensure it does turn out cheap. The agency hopes to award 'performance-based' contracts, so that the industry teams building the telescope get

for detecting gravity waves from space, next-generation orbiting X-ray observatories, and at least half a dozen approaches to space-based interferometry. Meanwhile, ESA is considering orbiting interferometers as possible 'cornerstone' missions after 2014, along with advanced observatories investigating the X-ray, gamma-ray and infrared regions of the spectrum to replace those launched in the next decade. But some astronomers are worried that some of the envisaged ESA missions may be preempted by NASA's — there would seem to be scope for more collaboration in planning even for the far future.

NASA has devoted much attention recently to plans for a large-scale search for planets beyond our Solar System (see *Nature*, 378, 650; 1995). The culmination of that programme would be an interferometry array for infrared spectroscopy located as far out, perhaps, as Jupiter. The scientific pay-off would be the ability to obtain spectra for the atmospheres of Earth-like planets, perhaps even finding evidence of life — the first attempt to use space-

craft to find such a signature on Earth was reported in *Nature* in 1993 (365, 715–721). The estimated date for that mission is 2015. ESA is studying a similar concept, called Darwin. Considering the expense of such a project, an international partnership would seem a good idea.

Financing two 'big things' at once is not likely in the current budgetary climate, so NASA proposes an intermediate step, a Space Interferometry Mission to test optical interferometers in Earth orbit. The price, says NASA's Ed Weiler, would be reasonable by space standards: \$300 million or so. And apart from testing the technology involved, the mission would have much to offer scientifically — microarc-second resolution, and the ability to measure true astrometric distances to any star in the Milky Way, including the Cepheid variable stars, critical to determining distances in the Universe.

All these programmes may eventually fly, if their costs can be contained. But the projects that get the nod must either be extraordinarily compelling or serve many customers. A good example of a worthy mission that fails this test, says one senior astronomer who serves on US review panels, is the Laser Interferometer Space Antenna gravity-wave detector — good science, but considering the expense, proportionately few researchers would benefit. The Next Big Thing will have to be more universally popular than that to survive. □



interferometry — as at the Very Large Array radio telescope near Socorro, New Mexico — is spacebound.

most of their money at the end, after they have produced, on time and within budget, an instrument that works as advertised. If the eight-metre design turns out not to be feasible for \$500 million, NASA would scale it back. "I certainly have no intention of raising the price," says Mather.

Scientists have an incentive to push the telescope's diameter as high as it can go, however. Seeing the earliest globular clusters of stars in the Universe — perhaps pre-dating the first galaxies — is among the prizes that an eight-metre instrument, but probably not a four-metre one, could claim, says Mather.

If the NGST is built, it is likely to be an international venture. The European Space Agency is keen on the idea, and wants to come in with a higher share than the 15 per cent it contributed to Hubble. The memorandum covering that particular partnership expires in 2001, but ESA will continue to provide personnel at the Space Telescope Science Institute in Baltimore as a gesture of goodwill. ESA is also considering paying part of the cost of re-boosting the Hubble in 1999 to maintain its orbital lifetime, and may build advanced instruments for installation in Hubble after 2002.

The NGST is not the only contender for the title of 'next big thing', even within NASA. The agency is funding 28 study teams to explore ideas for astrophysics missions after the year 2000. The ideas come in all sizes, and include large projects like the Laser Interferometer Space Antenna

► The XMM project is also notable for its systematic planning for 'serendipity observations'. Any field of view of the telescope contains not only the target object but also many others. Ensuring that they will also be preserved is a strong priority at the mission's data centre at Leicester University.

Space-based astronomers have been far more generous and systematic than their ground-based counterparts in archiving data and making them widely available over the Internet. That trend is likely to continue, but this is a tempting area in which to skimp: when ESA cut its science budget earlier this year, data archiving was among the first casualties. That must limit future returns on mission investment.

Another constraint on 'value for money' of data comes from the science that underlies their interpretation. As the submillimetre waveband opens up, astronomers worry whether we know enough molecular physics to do justice to the spectra. Only about half of the spectral lines in the interstellar medium have been identified, and laboratory measurements and quantum mechanical calculations of molecular transitions between rotational energy levels are difficult. Similar concerns apply to the X-ray regime, where there are uncertainties, for example, over probabilities of transitions between energy levels within key ions such as iron. Support for appropriate research in molecular and atomic physics is called for.

In short, the practice of space-based astronomy needs to change to make the best use of new data, and shows signs of doing so. But the ultimate value will not be in the number of bits and bytes returned, or even archived, but in the scientific questions it answers. The satellite missions now in the queue — plus those to follow in the next decade (see box, previous page) — should measure up well in this regard. As already discussed, SIRTf and other infrared satellites will investigate the early history of cosmic matter by looking at young galaxies. In so doing, SIRTf should outperform even the largest mountain-top telescopes equipped with adaptive optics.

AXAF, XMM, ESA's Integral and other high-energy satellites will explore frequencies unavailable from the ground to investigate some of the hottest regions in the Universe, including the behaviour of matter near the massive black holes at the centres of active galaxies — so called because of the enormous amounts of radiation emitted as the central black hole accretes matter from its surroundings. Such observations will also provide a unique insight into space-time itself which, according to Einstein's theory of general relativity, becomes strongly distorted near a black hole.

Some of the most perplexing problems in astrophysics, like the nature of dark matter and gamma-ray bursters, will benefit from the wide range of spectral coverage from

space, since no one is yet sure where the important clues lie. That balance of coverage, say many astronomers, is perhaps the greatest value of the new fleet of satellites preparing to launch.

Patrick Thaddeus of the Harvard-Smithsonian Astrophysical Observatory, who chairs a new National Research Council study of scientific questions for space-based astronomy, emphasizes this need for balance, saying that it is very important not to be overly influenced by trends, in technology or science. Many discoveries come through serendipity, for which large, multi-user observatories exploring a new part of the spectrum are ideally suited.

Thaddeus cites an example. If astronomers in the 1950s had gone searching directly for neutron stars, they probably would have failed, he says. As it turns out, they were discovered through the back door of radioastronomy. Perhaps the wish-list of astronomical problems to be solved, either from space or from the ground, should therefore be kept short and revised frequently. And if by 2020, as astronomers in their least sober moments hope, we have indeed discovered life around another star, it will be goodbye to budgetary downsizing.

Tony Reichhardt (Washington), Alison Abbott (Munich) & David Swinbanks (Tokyo)

for ESSENTIAL research papers ... nature REPRINTS

These are an effective reference tool for all

- presentations
- conferences and exhibitions
- training sessions
- your own personal files

Nature reprints are printed in full colour or in black and white. You can use your company logo, typesetting or covers, ideal for presentations. Some articles appearing in **Nature** may not be reproduced without consent of the publisher, therefore they are the only legal way to obtain the articles you need.

To get your reprints from **Nature** please contact our reprints department:

Nature, Reprints Department,

Porters South, Crinan Street, London N1 9XW UK

Tel: +44 (0)171 833 4000 Fax: +44 (0) 171 843 4640

Nature, Reprints Department,

345 Park Avenue South, 10th Floor,
New York, NY 10010 USA

Tel: (212) 726-9278 Fax: (212) 696-9591

Nature, Reprints Department,

Shin-Mitsuke Bldg., 36 Ichigaya Tamachi,
Shinjuku-ku, Tokyo 162, Japan

Tel: +81 (0)3 3267 8751 Fax: +81 (0)3 3267 8746

The case for the passive voice

SIR — Maddox's article¹ accusing scientists of being poor writers engendered a reply from one of the growing number of adherents to the use of the active voice in scientific writing². This practice appears to have arisen in the United States over the past 20 years or so and is now encouraged by authors of otherwise excellent texts on scientific writing. Day for example³. It is claimed that the use of the active voice encourages clearer and more exciting writing and that the use of the passive voice is more difficult for the reader and is an expression of false modesty⁴. The truth is that both the teaching of the English language and the standards expected of students have declined from the previous high standards upheld by educators in the English-speaking world. Simply put, the writing of precise prose in the passive voice has become too difficult for many of today's scientists. This is unfortunate for a number of reasons.

Using the passive voice in scientific writing allows the researcher to stand at a distance from his or her work. By standing at a distance, an unbiased viewpoint is much more likely to be reached. An unbiased viewpoint encourages a world view and an open mind, surely prerequisites for good science. Many scientific papers published today refer only to literature published in the past five years (in other words, easily located using one of the computer databases available), are parochial in nature and in many cases put forward old arguments as new. John Lawton terms this habit 'reinventing the wheel' and implies a 30-year cycle⁴.

The use of the passive voice encourages disciplined writing, cases must agree, tenses must be used correctly. It is therefore more demanding, but the precision and professionalism displayed is worth the effort. It is possible to be enthusiastic and to write stimulating and exciting prose using the passive voice. How many of the most memorable prose passages in English literature are written in the active voice?

Using the active voice is an easy option. There is no need to discipline one's thoughts. An author can just pour out his or her thoughts. This leads to careless presentation, particularly in methods and materials sections. As any editor knows, many papers submitted for publication appear to be first, or at the best second, drafts. Most authors using the active voice show no consistency of use. Papers alternate between passive voice statements and active voice statements, sometimes in the same paragraph, with no logic for the change of voice. This results in a paper full of inconsistencies and, of course, a general mixture of styles.

Using the active voice engenders possessiveness in the results and/or work. By engendering possessiveness an author risks adopting a biased and partisan stance. Wearing blinkers is no way to conduct good science. The active voice, with its less professional approach and tendency to foster the use of colloquialisms ("hassle" is one example I came across in a submitted manuscript) can make the writing appear quaint and amateur and akin to the offerings seen in amateur journals of natural history.

It is tempting when writing from a partisan viewpoint to descend to spite and denigration of other work. This too often manifests itself in biased and anonymous peer review of manuscripts and grants. There is also the possibility that use of the active voice and the resulting adoption of results and hypotheses as the author's own personal property leads to an unwillingness to see those results contradicted or refuted. This may, in the worst-case scenario, lead to the fabrication of results, something seen much more today than 20 years ago when the passive voice was *de rigueur*, as judged by the number of articles concerning the subject seen recently⁵⁻⁷.

In conclusion, the use of the passive voice encourages precision and probity, and when used correctly can generate as much passion and stimulation as the skilled use of the active voice. The active voice encourages carelessness, partisanship and, as used by many of its adherents, does no favours to the English language or science.

Simon R. Leather
Department of Biology,
Imperial College,
Silwood Park, Ascot,
Berkshire, SL5 7PY, UK

1. Maddox, J. *Nature* **378**, 521-533 (1995).
2. Avery, L. *Nature* **379**, 293 (1996).
3. Day, R. A. *Scientific English*, 2nd edn (Oryx, Arizona, USA, 1995); *How to Write & Publish a Scientific paper*, 4th edn (Cambridge Univ. Press, Cambridge, 1995).
4. Lawton, J. H. *Oikos* **60**, 273-274 (1991).
5. James, W. H. *Nature* **377**, 474 (1995).
6. Kerr, D. J. R. *Coll. Phys. Lond.* **29**, 461 (1995).
7. Lowry, F. *Can. med. Ass. J.* **152**, 1507-1508 (1995).

Questionable judgement

SIR — If your recent news report (*Nature* **380**, 371; 1996) is accurate, I find it very disturbing that a Brazilian court would rule, upheld upon appeal, as plagiaristic a 1991 publication describing results similar to unpublished data reported by another group at a 1982 meeting. The acknowledgement of the original group in the 1991 paper constituted as much credit as could reasonably be expected in any circumstance let alone one in which the data were not

published for more than eight years.

The Brazilian court system erred in permitting an investigator to claim some sort of intellectual property right over unpublished, not peer-reviewed and apparently not-publishable experiments. If courts elsewhere were to follow this precedent, scientific progress would be curtailed by dampening scientists' free incorporation of new concepts and the rapid confirmation of results afforded by this practice.

You should publish the address of the court so that scientists may encourage the magistrates to reverse their decision.

Fred Schaefele
Metabolic Research Unit,
University of California, San Francisco,
California 94194-0540, USA

■ Anybody wishing to do so should write, quoting Processo no. 253723-1, to Tribunal de Justiça do Estado de São Paulo, 4a Câmara de Direito Privado, Palácio da Justiça, Praça da Sé sem número, São Paulo — SP, 01001-001 Brasil □

Human rights in Belarus

SIR — 21 May 1996 was the seventy-fifth anniversary of the birth of Andrei Sakharov, the physicist and winner of the Nobel prize for peace.

But while scientists from around the world were gathering in Moscow to celebrate the anniversary, a few hundred kilometres to the west, in Minsk, Belarus, two scientists had been arrested for participating in a peaceful citizens' rally to mark the tenth anniversary of the Chernobyl nuclear accident.

Yuri Khadyka and Viachaslau Siuchyuk, who have followed Sakharov's example of combining scientific research with concern for human rights, are threatened with sentences of up to three years' imprisonment for alleged public order offences. In protest against their arrest, Khadyka and Siuchyuk have been on hunger strike, and the health and lives of both are threatened.

This anniversary is an appropriate time to make people elsewhere aware of the threat to human rights posed by the present regime in Belarus.

Alexander Lutzko
International Sakharov
College on Radioecology,
Dolgobrodskaya 23,
Minsk, Belarus

corres@nature.com

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only, or e-mailed (with full mailing address) to corres@nature.com

Pioneers of vaccination

SIR — In his praise of William Woodville for his role in vaccination¹, G. C. Cook is unjust to the London physician George Pearson (1751–1828). Pearson helped to introduce vaccination in Austria in 1799, and later indirectly into the Ottoman Empire and into India. The *Dictionary of National Biography*² mentions that “Pearson sent out letters to doctors in England and abroad with regard to his work” (on smallpox) but it is not widely known that he also forwarded cowpox-lymph. He impregnated threads with “cowpox matter” taken on the seventh to twelfth day from the pustules of vaccinated people, and fixed these threads on paper by means of wafers or sealing-wax. It was Pearson who invented this way of carrying vaccine.

In a letter dated 20 March 1799 he sent some of these impregnated threads to Dr Jean Peschier (1774–1831), a young physician from Geneva who was living in Vienna at that time^{3,4}. The only known connection between the two is that both had obtained their doctor's degree at the University of Edinburgh, Pearson in 1773 and Peschier in 1797.

Peschier offered Pearson's threads to the “Sanitätsreferent” (medical chief) of Lower Austria (which at this time included Vienna), Dr Pascal Joseph Ferro

(1753–1809). On 30 April 1799, for the first time outside England, Ferro vaccinated three of his own children, two girls and a boy, in Vienna by making shallow incisions in both upper arms, inserting a thread moistened with steam, and fixing it with an adhesive plaster^{4,5}. On 10 May 1799, the 3-year-old son of Dr Jean de Carro (1770–1857), a Viennese physician born in Geneva who also had a doctor's degree from the University of Edinburgh (1793), was inoculated by Ferro with pustule lymph from his younger daughter. On 20 May 1799, Carro transferred the “cowpox-pus” from this boy to his 18-month-old second son. On 15 July 1799 he inoculated both boys with true smallpox in the presence of several Austrian and foreign physicians, and they remained healthy and without symptoms^{3,4}. A few days later, Carro began to vaccinate children in Vienna and surroundings, and other physicians soon followed suit. Without Pearson's help there would have been no vaccinations in Austria at this early time because there were no cattle with cowpox.

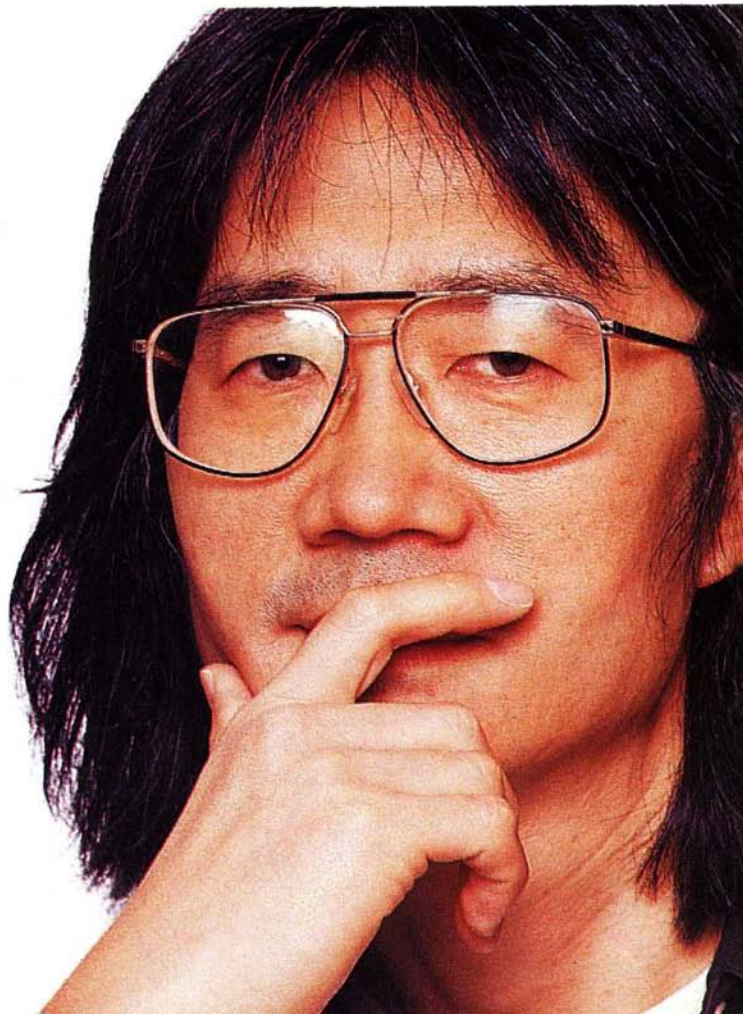
In the autumn of 1800, Carro (physician to Lord Minto, 1751–1814, British Ambassador at the court of Vienna) followed Pearson's lead and sent impregnated threads in small glass bottles to Lord Elgin

(1766–1841), then British Ambassador to the Ottoman Porte in Constantinople for the vaccination of his one-year-old son; Elgin wrote to Carro on 23 December 1800: “The vaccine of the arm of my son has served for the vaccination of several other children”. In February 1802, Carro despatched impregnated threads to the British Resident in Baghdad, Mr Harford Jones, and from there the vaccine was passed on to Bassora (Basra), Bombay, Calcutta, Madras, Goa, Sumatra, Ceylon and elsewhere⁶.

Jenner corresponded with Carro and on 27 November 1799 he wrote to him⁷: “Conscious of its importance it was always my hope that the subject would be taken up on the Continent and I am much gratified to see it fall into such able hands in Vienna”. On 30 March 1803 Jenner wrote that he had felt sad about the failure of his attempts to introduce vaccination into India. “Judge, therefore, the pleasure you have given me in telling me that my wishes are accomplished.”

In Vienna, there is a Jennerplatz, a Ferrogasse and a Carrogasse in remembrance of these three pioneers of vaccination. But George Pearson is also remembered in Vienna and Austria for his role in introducing vaccination against what was called by the Habsburg Empress Maria Theresia (1717–80) the “hereditary enemy of the imperial family”. It had killed her

To **save an hour**
each time you
purify histidine-
tagged proteins,
put it on the
tip of a syringe



uncle, Emperor Joseph I, three of her 16 children and the first and the second wives of her son, Emperor Joseph II; in 1767, she fell ill herself and barely survived, with her face badly pockmarked.

Friedrich Katscher

*Mariahilfer Strasse 133,
A-1150 Vienna, Austria*

1. Cook, G. C. *Nature* **381**, 18 (1996).
2. *Dictionary of National Biography* Vol. 44, 166 (1895).
3. De Carro, J. Lettre aux Redacteurs de la Bibliothèque Britannique, from Vienna, 27 July 1799. *Bibliothèque Britannique, Sciences et Arts* **11**, 337–346 (Geneva, Year VII, 1798–99).
4. De Carro, J. *Observations et expériences sur l'inoculation de la vaccine* (Kurtzbeck, Vienna, 1801); German translation: *Biobachtungen und Erfahrungen über die Impfung der Kuhpocke* (J. Geistinger, Vienna, 1801).
5. Ferro, P. J. *Erfahrungen über das Einimpfen mit Kuhblattereyter. Medizinisches Archiv von Wien und Oesterreich unter der Enns. Vom Jahre 1799* 2, 128–141 (F. J. Rötzel, Vienna, 1800).
6. De Carro, J. *Histoire de la vaccination en Turquie, en Grèce et aux Indes orientales* (J. Geistinger, Vienna, 1804).
7. Miller, G. *Letters of Edward Jenner*, 10 (Johns Hopkins Univ. Press, Baltimore and London, 1983).

School daze

SIR — In a News story about reviving science in UK schools (*Nature* **380**, 372; 1996), Ehsan Masood quotes Alan Smithers of Brunel University as asking: "How relevant is an 'A' level in biology for a better understanding of the BSE [bovine spongiform encephalopathy] crisis?"

After more than 30 years spent teaching

A-level Biology, I wish to answer his question. His basic tenet is that we should study 'relevant' subjects to prepare our students for the 'real' world. I am reminded of T. H. Huxley in the great evolution debate with Bishop Wilberforce in that, by choosing such a clear-cut example, he has exposed the weakness of his argument.

If we made our post-16 science teaching more specifically relevant, the students would fail to gain a sense of perspective and be ill-equipped to respond to any new idea (or in this case disease) because they would lack a context for it. Understanding depends on seeing the relationships between ideas. To study a series of current obsessions would be the worst way of gaining solid intellectual high ground.

This is largely what is wrong with the General Certificate of Secondary Education (GCSE). Instead of learning the role of microorganisms with a sideways look at applications such as sewage works, we study the sewage works with a sideways look at the organisms involved.

Not only is this terribly short-term — sewage works change — but it is also inherently boring. Intelligent people find fascinating the revelations about the nature of the world that science provides. Technology flows from this understanding. Using the principles of physics and mathematics, we may design cars for the future. Studying the current model without basic

science allows us only to make copies of it, and its study provides only limited intellectual interest.

After their GCSE experience of 'science', students are voting with their feet; they will desert in droves if Smithers has his way and A-levels become more of the same. I suggest the following solutions.

(1) Stop preprogrammed assessed 'practicals' which are knocking the spirit of genuine inquiry on the head. Students see them merely as a series of hoops to jump through to gain marks.

(2) Reward students who develop ideas or methods or carry out experiments on their own initiative to satisfy their curiosity.

(3) Put the principles of the subject first and refer only to the application so that students become aware of the relevance without having to learn it.

(4) Go as deeply into any topic as students can manage so as to stretch their minds and allow them to appreciate the complexity of the topic.

(5) Encourage teachers to respond to certain types of question from their students by discussing with the class how they might find out the answer experimentally and then doing the experiment, introducing a sense of spontaneity and panache.

J. R. Lloyd

*The Old Rectory,
Bradford Abbas,
Sherborne, Dorset DT9 6RF, UK*

Would you like to save one hour in histidine-tagged protein purification?
To load, purify and elute in 15 minutes? Now it's possible—time after time.

HisTrap kit: saving an hour couldn't be simpler

The HisTrap™ Kit presents you with a solution so simple it works with a syringe. Inside the kit is all you'll need for up to ten affinity purification runs: pre-packed cartridges, ready-made stock solutions, a protocol that's easy to follow, and, of course, a syringe.

You'll get high purity recombinant proteins as the kit lets you use the well-known binding of poly(histidine) sequences to immobilized nickel. What's more, it saves you the effort of weighing out buffer salts.

Molecular biologists have been purifying recombinant proteins with Pharmacia Biotech media from the very beginning. Call us at 1 (800) 526 3593 from the USA, +81 (0)3 3492 6949 from Japan or +46 (0)18 16 50 11 from Europe and the rest of the world (or meet us on the Internet at <http://www.biotech.pharmacia.se>). Ask about the kit that saves an hour.

**Pharmacia
Biotech**
Uppsala, Sweden. (And the rest of the world)

A column from
the HisTrap™ Kit,
magnified 280%.



"Purity, simplicity and time savings: a product that gives you all these, lets you take a break from the lab for a bagel, coffee and...hmm, maybe a haircut?" Kyungsun Suh, Ph.D., at the Rockefeller University, New York, NY

Beyond the Hubble Space Telescope

Harley A. Thronson Jr, Alan Dressler and Douglas Richstone

NASA needs to develop a new generation of space observatories embracing the theme of 'cosmic origins' if it is to maintain its lead in the exploration of space.

WHAT is past is prologue in astronomy: the content of the modern Universe — even the presence of life — is a product of events that took place billions of years ago. Fundamental questions about this epoch in cosmic evolution, about the origins of matter, structure and life in the cosmos, await a new generation of space observatories.

In 1993, the US National Aeronautics and Space Administration (NASA) and the Associated Universities for Research in Astronomy recognized that there were profound scientific questions that could not be answered with current space astronomy missions. Moreover, no new advanced space observatories were being considered for the period after the Hubble Space Telescope completes its primary science mission in 2005. In response, a committee of 18 astronomers from North America and Europe was established to identify core scientific questions about the cosmos that were unlikely to be answered with existing observatories and to recommend space missions capable of investigating them. This "HST and Beyond" committee, of which we are members, proposed "cosmic origins" as a unifying theme for a handful of fundamental problems involving the birth of normal galaxies, star formation and the possibility of Earth-like planets beyond our Solar System.

Fortunately, we can see the history of our Universe directly: the farther we peer out into space, the further back we see into time. But the first few billion years of cosmic history are shrouded from the view of even the largest telescopes on Earth. The difficulty arises because the expansion of the Universe causes light from distant — and, consequently, young — galaxies to be shifted to longer wavelengths. Light from normal galaxies, such as our own Milky Way, is emitted by billions of stars at wavelengths mainly between 0.4 and 2 micrometres. But the large redshifts of young galaxies mean that this ancient light reaches Earth shifted to a wavelength longer than about 2.5 micrometres. At these wavelengths, Earth's atmosphere becomes opaque, and what light struggles through is overwhelmed by the thermal emission from both the atmosphere and the telescope. Only the most luminous, and most atypical, young galaxies can be detected by Earth-bound telescopes: the birth of the most common galaxies are accessible only from space.

As well as being interested in the birth of stars and galaxies, scientists and the public are intensely interested in whether Earth is alone in the cosmos in supporting life. A search for the origins of life, beginning with a search for Earth-like planets, requires a technology different from that used to study *primaeval* galaxies: spatial interferometry, where the optical system of many different telescopes can be combined destructively to cancel out the blinding light from a central star. This leaves behind the faint but detectable emission from surrounding planets. This residual light would be searched spectroscopically at infrared wavelengths for signatures of molecules believed to be necessary for life.

Important steps in understanding our cosmic origins are already being taken by both NASA and the European Space Agency. The latter is operating the Infrared Space Observatory, which is targeting luminous, very energetic galaxies. Similarly, NASA will launch its own powerful Space Infrared Telescope Facility, which will probe into the history of the Universe. Both observatories are essential milestones, but both are limited by modest apertures and lifetimes.

As a result of these considerations, our committee last week published three recommendations to NASA intended to guide the agency in space science at the start of the next century. First, NASA should develop an infrared-optimized space observatory with an aperture of more than 4 metres to study normal galaxies in the early Universe. The wavelength of optimum sensitivity of this observatory — about 1–5 micrometres — covers diagnostic spectral features from a wide range of other objects: the youngest and oldest stars, Solar System objects and galaxies throughout the local Universe. Furthermore, enlarging the wavelength coverage of the facility would greatly increase its usefulness to the broader astronomical community.

Second, NASA should develop the capability for space interferometry at optical and infrared wavelengths. The next major direction in space astronomy will be a notable increase in angular resolution: that is, we will be able to study very fine structure and determine positions and motions extremely accurately. A challenging application of interferometry in space will be the mission to search for Earth-like planets

around neighbouring stars, and to explore their atmospheres for the molecules essential for life. The first main step in interferometry will be the Space Interferometry Mission, which will measure stellar positions to the limits of accuracy allowed by today's technology.

Third, the Hubble Space Telescope should continue to operate after 2005, when it is scheduled to complete its primary scientific mission, but in a 'no repair, no upgrade' mode. This would ensure the maintenance of a unique facility for ultraviolet astronomy and a continued return on the unprecedented US investment of thousands of man-years and billions of dollars in the observatory, whose success has dazzled people everywhere.

These three initiatives, along with a suite of smaller, more focused missions, will become the backbone of NASA's "Origins Program", which is intended to be a principal theme for future US research in space. In making these recommendations for the next generation of major ultraviolet–optical–infrared space astronomy missions, our committee was aware that NASA is likely to face severe constraints on its resources. So total costs cannot be comparable to those of the Hubble telescope. Alternatives to expanding costs range from applying innovative technologies, perhaps developed within defence-related industries, to exploring new management techniques, which would cut costs through accelerated development and international collaboration.

The initiatives recommended above would ensure NASA's continuing lead in the exploration of space well into the next century. By embracing our age-old human fascination with cosmic origins, such programmes will, we hope, excite humanity's imagination and encourage wider public participation in the exploration of the Universe. The recommendations do not chart an easy course for NASA, but the rewards of pursuing them are great. □

Harley A. Thronson Jr is in the Department of Physics and Astronomy, University of Wyoming, PO Box 3905, Laramie, Wyoming 82071, USA. Alan Dressler is at the Carnegie Observatories, 813 Santa Barbara St, Pasadena, California 91101, USA. Douglas Richstone is in the Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA.

Strengths and weaknesses in memory

Charles F. Stevens

How memories are formed is an enduring puzzle. Two studies further implicate a process by which synapses in the brain are made weaker, as well as stronger, and provide pointers as to how the two phenomena are regulated.

A WIDELY held belief among neurobiologists — although one without much direct experimental support — is that memories are stored as patterns of synaptic strengths. Two papers from groups led by Mark Bear, one which appeared last month¹, the other on page 526 of this issue², shed light on how synaptic strength is modified by experience and thus on how memories might be formed.

First, a little more detail about this idea of how a memory is represented. When a nerve impulse in one neuron arrives at a synapse, it sets into motion a complicated chain of events, one of which is the release of the neurotransmitter glutamate from the 'sending' side of the synapses we are considering. The result is a brief voltage change (a 'postsynaptic potential') in the target neuron, and the average size of this voltage change defines the synapse's strength. As might be expected for such a vital process as synaptic transmission, the brain has many ways for adjusting the strength of its synapses, and the study of these modulatory processes has preoccupied many neurobiologists over the past decade. A cubic millimetre of cortex contains about a billion synapses, so if each synapse could be either strong or weak, then that volume of cortex could store something like 100 megabytes of information. This number cannot be taken seriously for many reasons, but it does indicate the potential power, and thus the great attraction, of the notion that memories can be stored as patterns of synaptic strengths.

Synaptic strength can be modified in many ways, but the particular type of modification that has attracted the most attention is LTP (long-term potentiation), because these changes in strength last for days or even weeks, an obviously desirable quality for a memory mechanism. When a synapse is used repeatedly (in just the right way, as described later), it becomes much stronger, perhaps twice as strong, and stays that way: this is LTP.

On the principle of 'what goes up must come down', one might expect there to be an inverse process that can make synapses weaker. This inverse of LTP is called LTD (long-term depression), and its existence has been generally accepted for only about the past five years³. Indeed, the potential role of LTD as a memory mechanism, rather than as an experimental

curiosity seen only under restricted and artificial experimental conditions (slices of hippocampus from immature animals), has been up in the air because LTD had not been found in whole animal brains. This is where the first paper comes in.

Heynen *et al.*¹ used adult rats to carry out the same procedure *in vivo* that leads to LTD in brain slices and they found LTD with the same properties as seen *in vitro*.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Memorie, as represented by a man writing (engraved by T. Jenner in about 1650). The accompanying verse reveals a doleful view of the faculty: *A comon Inne all comers to retyne/A sive where good rune out & bad remayne/A burrow with a thousand vermin hydes/A den where nothinge that is good abides.*

Furthermore, this LTD saturates (synaptic strength reaches a limit beyond which it cannot go) as does LTP, and LTP and LTD can be used to erase one another at the same synapses. Because the LTD remained unchanged for the duration of the experiments, sometimes longer than five hours, these data seem to place LTD on an equal footing with LTP as a possible mechanism underlying memory. This certainly makes sense: the brain would want to store memories by adjusting synaptic strength both up and down to make this storage mechanism most flexible and efficient.

To appreciate the second paper, that by Kirkwood *et al.*², we need to delve a little into the intricacies of how LTP and LTD are produced. The puzzling thing is that the trigger for both mechanisms is firmly established to be the inflow of calcium ions through a particular type of ion chan-

nel present in the postsynaptic membrane, the famous glutamate-activated *N*-methyl-D-aspartate receptor (NMDAR) channel: to get either LTP or LTD, you have to use the synapse in the way that opens NMDAR channels and lets calcium flow into the cell. If an increase in calcium concentration near the postsynaptic membrane is the trigger for both LTP and LTD, what determines which one you get?

We know from experiments that using the synapses rapidly and repeatedly produces LTP (for example, 20 pulses at 100 Hz) and that slow, prolonged use (900 pulses at 1 per second, for example) results in LTD. The first guess might be that the key is the concentration of calcium near the postsynaptic membrane (high for LTP, low for LTD), but other work⁴ published earlier this year casts doubt on this idea. Although we still do not understand the precise conditions for getting LTP as opposed to LTD, Kirkwood *et al.*² start to define some of the relevant variables.

If a pattern of synaptic strengths is to provide a representation of memory, and if LTP and LTD are used to modify synaptic strengths, then one important question clearly is how the decision is made whether a particular synapse is to increase or decrease its strength. The work of Kirkwood *et al.* was motivated by the theoretical studies, published by Leon Cooper and colleagues⁵, that consider this question. This theory proposes that synapses should remember how much they have been used and should favour LTD with a lot of past use and LTP with little past use.

Kirkwood *et al.* therefore examined LTD and LTP in visual cortical slices from rats that had been reared normally or ones that had grown up in the dark. The guess is that LTD would be larger in normal rats than in dark-reared rats because the synapses in normal rats would remember having been used a lot, whereas the synapses in visual cortex from dark-reared rats would have seen much less use. And this is what they found. The nearly absent LTD observed in dark-reared rats was not just a general effect on LTD throughout the brain because hippocampal LTD was the same in normal and dark-reared rats.

If the history of use biases synapses towards LTP or LTD, how long do the synapses remember the extent of their previous use? By putting the dark-reared rats in the light for various periods of

Wellcome Institute Library, London

time, Kirkwood *et al.* found that LTD in dark-reared rats was back to normal after two days in the light. Further, they studied the change in synaptic strength as a function of stimulation frequency used to produce it and found LTD results below some critical frequency and LTP above that frequency; the greater the stimulation frequency, the larger the LTP. This critical frequency is shifted along the frequency axis by dark rearing so that LTP results at lower frequencies in the dark-reared than in normal rats.

From these two papers^{1,2} it seems that LTD could serve as a genuine inverse of the better studied LTP in the long-term regulation of synaptic strength, and we have a clue about how the threshold might be set to determine if stimulation will give LTP or LTD. Two pressing needs are

evident: neurobiologists must dig deeper to uncover the underlying cellular and molecular mechanisms, and they must investigate LTP and LTD at higher levels of organization to decide if these phenomena really can explain some kinds of memory. □

Charles F. Stevens is in the Laboratory for Molecular Neurobiology, The Salk Institute for Biological Studies, 10010 N. Torrey Road, San Diego, California 92186-5800, USA.

1. Heynen, A. J., Abraham, W. C. & Bear, M. F. *Nature* **381**, 163–166 (1996).
2. Kirkwood, A., Rioult, M. G. & Bear, M. F. *Nature* **381**, 526–528 (1996).
3. Bear, M. F. & Malenka, R. C. *Curr. Opin. Neurobiol.* **4**, 389–399 (1994).
4. Neveu, D. & Zucker, R. S. *Neuron* **16**, 619–629 (1996).
5. Bienenstock, E. L., Cooper, L. N. & Munro, P. W. *J. Neurosci.* **2**, 32–48 (1982).

X-RAY ASTRONOMY

Shrill notes from the stars

Guy Miller

ASTRONOMERS gather most of their information about the cosmos through spectral analysis and the interpretation of images. So it is notable that at a meeting last month in San Diego* the most exciting announcements involved neither images nor spectra. Instead, the highlight of the conference was the discovery of strange, extremely rapid variations in the brightnesses of several Galactic X-ray sources.

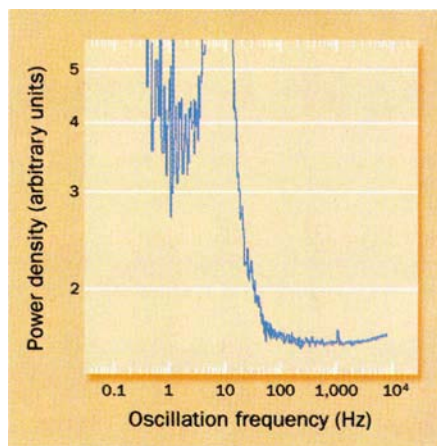
The sources in question are thought to be systems in which a weakly magnetized neutron star and a normal star orbit one another. The neutron star has a mass comparable to that of our Sun but is only a few tens of kilometres across, so the gravitational field near its surface is extremely strong. When gas lost by the companion star falls onto the neutron star, its tremendous gravitational potential energy is dissipated, and most of it goes into X-rays.

The gas has too much angular momentum to fall directly onto the neutron star, but instead spirals in slowly through a flattened accretion disk. If the magnetic field is very weak, the disk flow extends down to the stellar surface, ending in a hot, strongly sheared boundary layer. If the field is strong enough, the disk does not reach the stellar surface, but ends when magnetic stresses disrupt it at the magnetospheric boundary; material completes its journey to the stellar surface by moving along field lines to the star's magnetic poles.

The details of the flow structure and radiation mechanisms are unknown. Also unknown is the interior structure of the neutron star, where matter is crushed to densities higher than are found anywhere

else in the Universe (excluding, perhaps, the unobservable end state of material trapped by a black hole).

The Rossi X-ray Timing Explorer (RXTE) satellite, launched last December, has provided astronomers with a powerful tool for studying rapid X-ray variability, which should be an excellent indicator of conditions near neutron stars. A large proportional counter enables RXTE to collect and record X-ray photons at unprecedentedly high rates, so it



Variability of the X-ray binary Scorpius X-1, over a wide range of frequencies. A startling quasi-periodic oscillation at 1,100 Hz (the small peak on the right) has been seen by NASA's new satellite, RXTE.

can be used to study variations at higher frequencies and in greater detail than ever before. Not surprisingly, the first results from RXTE contained surprises.

Observations of the rapidly accreting source Scorpius X-1 (M. van der Klis, Univ. Amsterdam) show two new

high-frequency quasi-periodic oscillations (QPO). A QPO is a narrow peak in the power spectrum of source variability — not a perfect 'tone' but a note spread over a small range of frequencies. Previous examples of QPO typically had frequencies lower than 100 Hz (ref. 1).

The new QPO are different. One is at about 800 Hz, and appears only at lower accretion rates; the other has a frequency centroid that rises from 1,060 to 1,130 Hz with increasing accretion rate. Both have amplitudes of less than about 1% of the total emission. Their reported widths, tens of Hz, are probably due to changes in frequency during the observations — the true widths are probably much smaller. Strangely, the frequency of the 1,100 Hz QPO seems to correlate linearly with the frequency of a previously known QPO that lies between 6 and 20 Hz.

The hoped-for periodic oscillation due to the rotation of the neutron star was nowhere to be seen, but evidence for stellar rotation may be present in the more slowly accreting source, 4U1728–34, which was observed during a recent series of X-ray outbursts (T. Strohmayer, NASA GSFC). Early in each outburst, as the luminosity of the source was still rising, a very strong QPO was seen at 363 Hz. In two of the outbursts it had a width of about 0.5 Hz. Later, as the X-ray luminosity declined, the width of the QPO peak became immeasurably small. Two other oscillations were present, at about 800 Hz and 1,100 Hz with a variable separation of roughly 363 Hz.

Kilohertz QPO have been discovered in two other sources (W. Zhang, NASA GSFC; I. Lapidus, Univ. Cambridge), but so far all these discoveries pose more questions than they answer. Weakly magnetized accreting neutron stars are expected to have rotation periods of 1 to 10 ms, and to be the progenitors of millisecond radio pulsars. But only in 4U1728–34 has a periodic oscillation been found that can be identified with stellar rotation. Even there, episodes occur during which the 363 Hz oscillation loses coherence and is broadened by 0.5 Hz. A repetitive modulation of the X-ray emission on a two-second timescale would provide the necessary broadening, but it is not evident why that should occur.

Even more mysterious are the variable-frequency QPO. Models of these must account simultaneously for excursions in frequency by 10% or more, and at the same time for the narrowness of the features. The oscillation mode must be sensitive enough to conditions in the accretion flow or in the outer layers of the neutron star to vary appreciably, and yet not be so sensitive that transient, localized irregularities alter its frequency too much and destroy its high-frequency coherence. The mode must also be able to produce intensity variations of 1 to 10%.

*Meeting of the AAS High Energy Astrophysics Division, San Diego, California, USA, 29 April–4 May 1996.

A number of candidate mechanisms wait in the wings. If the oscillation originates in the disk flow, it must come from near the neutron star, where the orbital and hydrodynamic timescales are short. The inner edge of the accretion disk, whether touching the stellar surface or kept away from it by the stellar magnetic field, is a natural site for the production of kilohertz variability via hydrodynamic oscillations. In sources such as Cygnus X-2, a class of QPO exists that is widely believed to arise from the interaction of the disk flow with the magnetosphere^{2,3}. Searches for relations between this activity and kilohertz QPO will help in showing whether the latter have any connection with the inner edge of the disk.

Another possibility is that accretion or thermonuclear burning of the accreted material⁴ excites oscillatory modes in the surface layers of the neutron star. Changes in the thermal state of the surface could be invoked to explain the frequency excursions of the observed oscillations. At present, however, it is not clear that stellar oscillations can achieve the necessary frequency variations or drive variations in the X-ray output as large as those observed.

A final proposal made at the conference (R. I. Klein, Univ. California, Berkeley) is that the accreting hot, magnetized plasma at the stellar surface is subject to a radiation-hydrodynamic instability that causes photons to collect in 'bubbles'. The X-ray intensity should rise briefly each time a photon bubble escapes from the accreting gas. Computer simulations show that under some conditions the radiative frothing of the accreting gas becomes organized into collective photon-bubble oscillations⁵.

Although some mechanisms appear to have more difficulties than others, no leading candidate has emerged. It is not even certain whether there is a sole mechanism underlying the oscillations or a variety of causes. Nevertheless, the existence of extremely well-defined variabilities in the radiation from accreting neutron stars suggests that a powerful observational probe of these X-ray producing engines is at hand. It will depend on the development, in tandem, of a complete observational phenomenology and a secure theoretical framework. Clearly, many surprises lie ahead. □

Guy Miller is at the Dearborn Observatory, Northwestern University, 2131 Sheridan Road, Evanston, Illinois 60218, USA.

Snapping social swimmers

Jon Seger and Nancy A. Moran

THE ants of Death Valley (elevation -86 m) remain the lowest expression of eusociality on Earth, but on page 512 of this issue¹ J. E. Duffy describes eusocial colony organization in a snapping shrimp, *Synalpheus regalis*, which takes "the apex of animal social organization" to depths of a novel kind. This remarkable discovery establishes two new beachheads for eusociality: one is geographical (the oceans), and the other is taxonomic (class Crustacea, where eusociality was predicted to occur², but where it remained unknown before the appearance of Duffy's report).

Colonies of *S. regalis* occupy the internal canals of sponges on Caribbean coral reefs, each colony containing just one reproductive female. Most colony members appear to be closely related to the reproductive female (probably as offspring), and they appear to function as non-reproductive workers that defend their sponge against potential usurpers. So in all important respects the social organization of *S. regalis* resembles that of many eusocial terrestrial animals.

Eusociality has played an important role in the development of evolutionary biology. Darwin was famously troubled by worker ants because they posed a compound problem for the theory of natural selection: how could sterility itself evolve, and how could a sterile caste evolve novel morphologies and behaviours? In Darwin's time the known social-insect species seemed almost qualitatively different from their solitary relatives. Today, eusocial lifestyles are known in a remarkable diversity of animal taxa and there is no clear zone of demarcation between highly 'advanced' social insects with large colonies and elaborate caste systems, and the thousands of relatively inconspicuous 'subsocial' species that cooperate in one way or another in the protection or feeding of offspring. The three conventional pillars of eusociality are cooperative brood care, reproductive division of labour and overlap of generations³⁻⁵. It now seems that almost all of the conceivable variations on these themes will be found to occur somewhere in the lengthening list of social taxa.

As our knowledge of the diversity of social taxa has grown, the problem of sociality has become one of description as well as of explanation^{6,7}. What are the relevant

dimensions on which social systems vary? This is not as obvious as it seemed when fewer examples were known, and when they were known in less detail. As the variation among social systems comes to seem more a matter of quantities than of qualities, the search for satisfying and testable causal hypotheses also becomes more difficult. What set of forces carries a diverse set of lineages to a particular region of 'social-system space'? The discovery of social snapping shrimp will enliven



J. E. Duffy

A coral-reef snapping shrimp (*Synalpheus regalis*) that lives in eusocial colonies¹. The photograph shows part of a colony after removal from the host sponge. The sole reproductive female ('queen') is at the centre, and several large colony members display the distinctive fighting claw that they use to defend the colony's sponge.

discussion of these issues, and it will surely stimulate searches for other taxonomically isolated occurrences of eusociality.

So why did these shrimp become social? Duffy's report stresses several factors that link many origins of sociality, including the cavity-nesting habit, direct development and the possession of a formidable weapon, in this case the powerful 'snapping' major chela, or fighting claw, which is undoubtedly as effective as the hymenopteran sting in defending a nest and its brood. In snapping shrimp, males as well as females are diploid, in contrast to the Hymenoptera, where males are haploid. Modern thinking about the evolution of sociality begins with W. D. Hamilton's observation that the unusually close genetic relatedness of full sisters in haplodiploid species ($r = 3/4$, rather than $r = 1/2$) might, under certain circumstances, induce females to help their mothers rear more siblings, as an alternative to producing offspring of their own^{8,9}. Snapping shrimp, naked mole-rats, termites and other diploid social taxa show that asymmetrical coefficients of relatedness are not essential, even if they are sometimes helpful.

Since 1977, eusociality has been docu-

1. van der Kils, M. A. *Rev. Astr. Astrophys.* **27**, 517-533 (1989).
2. Alpar, M. A. & Shaham, J. *Nature* **316**, 239-241 (1985).
3. Lamb, F. K., Shibazaki, N., Alpar, M. A. & Shaham, J. *Nature* **317**, 681-687 (1985).
4. Strohmayer, T. & Lee, U. *Astrophys. J.* (in the press).
5. Klein, R. I., Arons, J., Jernigan, G. & Hsu, J. J.-L. *Astrophys. J.* **457**, L85-L89 (1996).

mented in several groups of gall-making aphids^{10,11}, in gall-making thrips¹², and now in sponge-dwelling crustaceans. This high rate of discovery implies that many more cavity-nesting social taxa remain to be found, and it supports the view that cavities promote sociality in large part by keeping relatives together and thereby creating opportunities for kin-selected reproductive altruism. These recently discovered cases are ones in which the cavity provides both shelter and food. We suspect that this association will turn out to be part of a relatively distinctive and common social syndrome. *Synalpheus* appear to feed on particles delivered to their door by the host sponge's feeding current. In the more familiar examples of hymenopteran and vertebrate sociality, non-reproductive helpers spend much of their time away from the nest, foraging for food to be brought back to nest-bound offspring of the reproductives.

Cavity dwelling may predispose a lineage to experiment with sociality, but these recent additions to the inventory of social taxa (like many that have been known for a long time) all have close relatives with similar living arrangements that are not social. How does a diversity of social systems evolve among taxa that inherited nearly identical developmental and

ecological constraints from their common ancestors? As the list of independent derivations of sociality grows, it provides a richer basis for comparative and phylogenetic tests of hypotheses to explain why sociality does and does not evolve under circumstances that may appear to be very similar. These newly discovered instances (and others to follow) seem certain to deepen our understanding of the causes of sociality in ways that would not have been possible without them. □

Jon Seger is in the Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA. Nancy A. Moran is in the Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA.

1. Duffy, J. E. *Nature* **381**, 512–514 (1996).
2. Spanier, E., Cobb, S. J. & James, M.-J. *Oikos* **67**, 573–576 (1993).
3. Batra, S. W. T. *Indian J. Ent.* **28**, 375–393 (1966).
4. Michener, C. D. *A. Rev. Ent.* **14**, 299–342 (1969).
5. Wilson, E. O. *The Insect Societies* (Belknap, Cambridge, MA, 1971).
6. Sherman, P. W., Lacey, E. A., Reeve, H. K. & Keller, L. *Behav. Ecol.* **6**, 102–108 (1995).
7. Crespi, B. J. & Yanega, D. *Behav. Ecol.* **6**, 109–115 (1995).
8. Hamilton, W. D. J. *theor. Biol.* **7**, 1–52 (1964).
9. Hamilton, W. D. A. *Rev. Ecol. Syst.* **3**, 193–232 (1972).
10. Aoki, S. *Kontyu (Tokyo)* **45**, 276–282 (1977).
11. Stern, D. L. & Foster, W. A. *Biol. Rev.* **71**, 27–79 (1996).
12. Crespi, B. J. *Nature* **359**, 724–726 (1992).

PLANETS

Looking for the twilight zone

David Black

INTEREST in the nature of other planetary systems has increased considerably over the past decade. One reason is that the only way that we will ever understand how our own system formed is to obtain information about others. Another reason is that planets are like cosmic petri dishes, in that they provide the environment necessary for the development of life. This combination has given rise to a number of efforts to assess the likelihood that other planetary systems exist and that they have planets with the 'right stuff' — the chemical and physical prerequisites for life. One vital factor is where planets form, and a surprising answer to that question has been given by G. W. Wetherill, publishing in *Icarus* earlier this year¹.

Dole was the first to examine^{2,3} what was known about the conditions necessary for the existence and evolution of life in the context of planetary formation and evolution. His approach was simplistic by modern standards, and much has since been learned of how planetary companions to stars are likely to form. One of the more comprehensive recent assessments in this area⁴ emphasizes the role of climate and atmosphere on the existence of life.

A number of researchers have devel-

oped sophisticated computational tools^{5,6} to study how the Sun's retinue of planets, particularly the terrestrial planets, formed by accumulation of objects ~ 0.1–10 km across called planetesimals, first into 'embryos' (objects about the size of the Moon) and finally into planets like Mars, Mercury, Venus and the Earth. The same techniques can reveal the effect of varying certain global parameters, such as the surface density of the disk (that is, density

HABITABLE ZONES DURING THE FIRST BILLION YEARS

Stellar mass (solar masses)	Habitable zone (AU)
0.5	0.2–0.4
1.0	0.82–1.40
1.5	1.70–2.80

projected onto the disk plane) or the mass of the central star.

A concept central to all discussions of life beyond the Solar System is the 'habitable zone' (HZ). This is a region of space, centred on the central star, whose inner and outer boundaries were originally defined by the distance at which water would boil and freeze respectively. The underlying premise here is that life is water

based, and it should be noted that surface conditions on planets that fall within the HZ should allow simple organisms such as bacteria to survive, but may not be compatible with the presence of more advanced life forms.

Early efforts to define the HZ simply made some assumption about the albedo of a planet, and then were able to define where the zone fell for a given spectral type of star. More realistic models³ reflect the fact that habitable planets will probably have dynamic atmospheres and climates. The boundaries of the HZ are then the locations where a runaway greenhouse effect would occur (inner) and where CO₂ clouds would increase a planet's albedo sufficiently to lower its temperature below the freezing point of water (outer). The HZ for stars of 0.5, 1.0 and 1.5 solar masses, valid for the first billion years in the evolution of a star, are given in the table (from ref. 4).

The other half of the question is where the planets form. Using a numerical code, Wetherill¹ has performed five hundred simulations of planetary formation in the inner regions of model systems. The study concentrates on four attributes of forming systems: the mass of the central star, the surface-density distribution of embryos at a distance of 1 AU (the distance from the Earth to the Sun), the radial dependence of the surface density of solids, and the position and presence of gas-giant planets.

With all other factors fixed, there is little dependence of the resulting planetary systems on stellar mass in the range considered (0.5 to 1.5 solar masses). Most systems have a few planets of mass comparable to that of Mercury or larger in the region roughly 0.5 to 3.0 AU from the star. The most massive planets tend to be located near 1 AU and have masses comparable to that of the Earth. There is a high probability of finding at least one planet in the HZ in the case of a solar-mass star, but generally there are only small (Mercury mass or so) planets in the HZ for stars of 0.5 or 1.5 solar masses.

If instead the total mass of the protoplanetary disk is proportional to the mass of the star, the simulations produce systems where the typical masses of terrestrial planets tend to mimic the trend in disk density. Their spatial distribution is essentially the same as before, but there are slightly more massive planets in the HZ around the 1.5-solar-mass stars than when surface density was independent of stellar mass. A similar dependence of planetary mass and location on surface density is found for fixed stellar mass.

All of the previous simulations assumed that planets the mass of Jupiter and Saturn were present at their actual Solar System locations. If that assumption is removed, large terrestrial-mass planets are formed at much greater distances — they are no longer kept from forming by

the strong gravitational interaction arising from the giant planets. This raises the number of planets in the HZ, particularly for more massive stars.

This is a step beyond the early efforts of Dole and others, but it is only an early exploration of the problem. As more observational evidence is gathered on the nature of disks around young stellar objects, this type of study will gain more focus and insight.

One immediate implication, however, is that around stars roughly similar to the Sun there is a reasonable chance of finding planets suitable for providing life with a place to start. The probability is not one, but neither is it a tenth of a per cent.

It should be noted that this expectation is only for stars that have planetary systems at all, and we have no observational evidence to tell us whether the majority of stars have such systems, or whether the Sun is unusual in that regard. Wetherill's study alone will not alter current tactics in the 'search for extraterrestrial intelligence' (SETI), but combined with the results from searches for planetary systems that will emerge in the years to come it could well affect the approach to SETI.

Another aspect of the new study is of interest. The past year has seen the discovery of substellar-mass companions to a handful of stars⁷⁻⁹. What has been surprising is that all of these companions are located close to their parent star — the companion to 51 Pegasi is just 0.05 AU away — and they are all of about one Jupiter mass or greater. As shown by Wetherill's study, this result is more than counter-intuitive: it is inconsistent with the basic concepts of our current understanding of planetary system formation. Did these companions form at great distance and migrate inwards¹⁰? That might account for one system, but it is unlikely that all of the newly discovered systems can be accounted for in this way. It is much more likely that these objects formed more or less where they are at present, and that they are not planets. Are they low-mass brown dwarfs, or are they a new class of object¹¹? □

David Black is at the Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston, Texas 77058, USA.

Foams flow by stick and slip

David A. Weitz

FROM the top of a beer to the meringue on a pie, from shaving cream to a bubble bath, we encounter foams everywhere. But consider their properties; they are fascinating materials. They are made only of gas and liquid, materials that flow freely, yet foams can be rigid solids like the whipped cream on a dessert. This solid-like behaviour results from the foam microstructure, gas bubbles in a fluid, with interfaces stabilized by surfactant. But foams also flow easily, as anyone savouring the whipped cream of the dessert will

most foams contain enough liquid that more general physical concepts must be used to describe their properties. Durian takes a completely different tack: he concentrates on the bubbles as entities, and models their behaviour by considering the interactions between neighbours, thus directly including the critical length scale that controls the dynamics and properties of a foam.

To investigate the flow of foams experimentally, Gopal and Durian used novel imaging methods which exploit multiply scattered light⁷ to probe the bulk of a foam flowing under an applied shear. They observed a pronounced change in the dynamics once the applied distortion exceeded unity, so that the bubbles were forced to move beyond their local neighbours. The foam flowed by random discrete rearrangements of neighbouring bubbles, distinguished by a characteristic rate constant, implying that the topological rearrangements were localized to a well-defined size.

Warner Bros (Courtesy Kobal)

IMAGE
UNAMIGABLE
FOR A COPYRIGHT
FOR REASON
REASONS

Ensign Pulver (Jack Lemmon) discusses his foam problems with two colleagues (Henry Fonda and William Powell). From *Mister Roberts* (1955).

affirm. What makes foams solid? How do they flow? Important new insight has come from a recent experiment¹ by A. D. Gopal and D. J. Durian and a computer simulation² by Durian. The results emphasize the critical importance of the disordered packing of the bubbles, and show that foam flow occurs through topological rearrangements of bubble positions, sharing many features with the highly non-linear dynamics of the avalanches observed in granular flow and the intermittent 'stick-slip' dynamics of earthquakes.

Most of the recent theoretical work on foams has concentrated on the behaviour of the thin film between the bubbles, because simple rules can be used to describe its properties³⁻⁶. The challenge has been to generalize these rules to the more complex microstructure of real foams. While progress has been achieved, many uncertainties persist, particularly for the three-dimensional packing of real foams. Furthermore, these simple rules are strictly applicable only to dry foams, with no liquid between the bubbles. In reality,

These experimental conclusions were supported by the simulations, which also looked directly at the behaviour of bubbles through their interactions with their nearest neighbours. This allowed Durian to analyse the local motion, at least in the two-dimensional packings studied. The intermittent stick-slip motion of the bubbles was observed both in their motion and in the resultant stress on the walls, which decreased sharply when an avalanche-like rearrangement of the neighbouring bubbles occurred. These rearrangements were highly localized — one central bubble moved the most, with the remainder of the motion restricted to a halo of near neighbours.

The localized rearrangements observed by Durian contrast with the predictions for ordered model foams³, where the rearrangements occur simultaneously throughout the sample. They also contrast with the predictions for two-dimensional dry foams, which exhibit long-range, power-law correlations, reminiscent of self-organized criticality^{6,8}. Such behaviour does not appear to be important for real foams, or even for two-dimensional model foams that are not dry.

The simulations also predict the elastic modulus of the foam and the osmotic pressure (the pressure that must be applied to deform the bubbles to achieve a given volume fraction of gas), and agree,

1. Wetherill, G. W. *Icarus* **119**, 219–238 (1996).
2. Dole, S. H. *Habitable Planets for Man* (Blaisdell, New York, 1964).
3. Dole, S. H. *Icarus* **13**, 494–508 (1970).
4. Kasting, J. F., Whitmire, D. P. & Reynolds, R. T. *Icarus* **101**, 108–128 (1996).
5. Greenberg, R. J., Wacker, J., Chapman, C. R. & Hartman, W. K. *Icarus* **35**, 1–26 (1978).
6. Wetherill, G. W. & Stewart, G. R. *Icarus* **77**, 350–357 (1989).
7. Mayor, F. & Queloz, D. *Nature* **378**, 355–359 (1995).
8. Marcy, G. W. & Butler, R. P. *Astrophys. J.* (in the press).
9. Butler, R. P. & Marcy, G. W. *Astrophys. J.* (in the press).
10. Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. *Nature* **380**, 606–607 (1996).
11. Black, D. C. *Sky & Telescope* (in the press).

at least qualitatively, with experiment.

The imaging of foam dynamics is limited by the high opacity of the foams, and Durian's simulations were restricted to two dimensions. This limitation has been overcome in a three-dimensional simulation by LaCasse *et al.*⁹, using a similar model. Although restricted to static behaviour, their results were in quantitative agreement with experiment, both for the elasticity and the osmotic pressure. While validating Durian's basic conceptual approach, they incorporated more microscopic information about the bubble interactions, showing that the model of harmonic springs adopted by Durian is modified by the non-local nature of the shape deformations.

The properties of foams have puzzled scientists for centuries. They are not only important for understanding the behaviour of cosmetics and foods, but are also crucial for designing improved foams for other technologically important uses, such as fire-fighting and a wide range of chemi-

cal separation processes. Durian's results give us a new perspective on these questions by concentrating on the length scale of the bubbles, rather than that of the films between them. That should lead to significant progress in this fascinating field. Moreover, the results may be applicable to many other granular or particulate systems. □

David A. Weitz is in the Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd Street, Philadelphia, Pennsylvania 19104-6396, USA.

1. Gopal, A. D. & Durian, D. J. *Phys. Rev. Lett.* **75**, 2610 (1995).
2. Durian, D. J. *Phys. Rev. Lett.* **75**, 4780 (1995).
3. Princen, H. M. J. *Coll. Inter. Sci.* **91**, 160 (1983).
4. Kraynik, A. M. A. *Rev. Fluid Mech.* **20**, 325 (1988).
5. Herdtle, T. & Aref, H. J. *Fluid Mech.* **241**, 233 (1992).
6. Hutzler, S., Weaire, D. & Bolton, F. *Phil. Mag. B* **71**, 277 (1995).
7. Durian, D. J., Weitz, D. A. & Pine, D. J. *Science* **252**, 686 (1991).
8. Okuzono, T. & Kawasaki, K. *Phys. Rev. E* **51**, 1246 (1995).
9. LaCasse, M.-D., Grest, G. S., Levine, D., Mason, T. G. & Weitz, D. A. *Phys. Rev. Lett.* **76**, 3448 (1996).

VOLCANIC ERUPTIONS

Turbulent times at Taupo

Tim Druitt

As silica-rich magma ascends from deep beneath a volcano, dissolved water and carbon dioxide come out of solution violently, fragmenting it into a mixture of hot particles and gas that erupts explosively at the surface. If this mixture is lighter than air it ascends buoyantly as an eruption plume, often high into the stratosphere. But if denser, it collapses back like a fountain and pours over the landscape as a ground-hugging 'pyroclastic flow'. On page 509 of this issue¹ Dade and Huppert present a new model of the transport and sedimentation of the particularly violent Taupo pyroclastic flow which erupted 1,800 years ago in New Zealand, and covered 20,000 km² of the North Island under a blanket of red-hot pumice and ash.

Some pyroclastic flows, those with

small volume, are hot, dense avalanches of rock and dust which travel as valley-confined tongues for up to several kilometres. Their deposits have the steep fronts, side levees and lobate morphologies typical of other highly concentrated geophysical dispersions such as mud flows and cold rock avalanches².

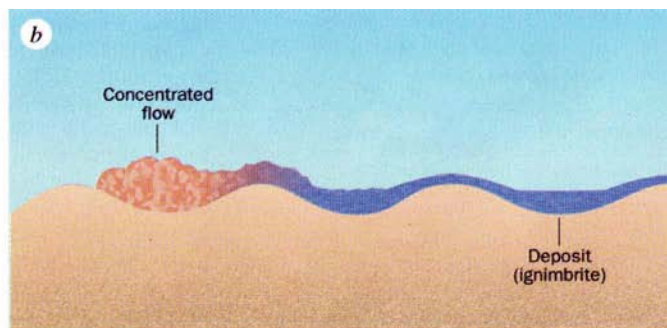
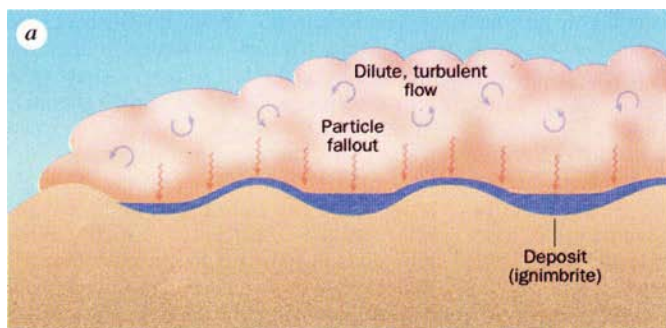
On a much larger scale, flows from single eruptions may have volumes of tens or hundreds of cubic kilometres. Their deposits, termed ignimbrites, must be very fluid at the moment they are laid down, because they form flat-topped ponds in valleys and depressions³. Large-scale ignimbrite eruptions are rare on a human timescale — even the 1991 eruption of Mount Pinatubo in the Philippines discharged only 5 to 7 km³ of ignimbrite.

The volcanic record has also yielded evidence of pyroclastic flows of exceptional violence. They produce low-aspect-ratio ignimbrites, which are relatively thin but spread very widely. One is the exceptionally well documented Taupo pyroclastic flow, which travelled 80 km in all directions from its erupting vent and surmounted obstacles over 1,500 m in elevation, implying a very high velocity⁴. The Taupo flow laid down a veneer of sediment all over the mountainous terrain, thickening in valleys and depressions into the flat-topped accumulations more typical of ignimbrite.

Since the mid-1970s, it has been believed that large pyroclastic flows, like their small-volume equivalents, travel as concentrated particle dispersions³. In this interpretation it is the high density of the flow that permits the transport of coarse volcanic debris over large distances. Ignimbrites are poorly sorted by size, which is attributed to the flows' having solid-concentrations of several tens of per cent, inhibiting segregation of particles of different sizes. The concentration of large, frothy pumice blocks at the tops of many ignimbrites is explained by the flotation of these low-density 'clasts' in the fluidized dispersion of ash, volcanic rock and gas that constitutes the flow. Laboratory fluidization experiments appear to support this model⁵.

However, some researchers have begun to question this picture, particularly for those low-aspect-ratio ignimbrites formed by fast flows. Valentine⁶ showed that a dilute, highly turbulent flow could transport coarse debris further than previously calculated. Perhaps some high-velocity pyroclastic flows travel as dilute suspensions in which intense turbulence replaces high concentration as the main support mechanism. In this model, depositional features such as pumice flotation are acquired at a late stage in the sedimentation process, and do not reflect the nature of the dilute transport system.

The dilute model gained some support from the volcanic blast that occurred on 18 May 1980, at Mount St Helens in



Two contrasting models for the emplacement of the Taupo ignimbrite. *a*, In the dilute model (refs 1 and 6) the flow particle concentration is very low. The sediment layer is at first highly fluidized by escaping gases and drains into topographic lows, leaving only thin veneers on ridge crests and valley sides. *b*, In the second model (ref. 4), the flow

is relatively thin and highly concentrated (tens of per cent), and the deposit is laid down by the tail of the flow. The dilute flow can traverse mountainous terrain readily because its thickness is greater than, or comparable to, the topographic roughness. The thin, highly concentrated flow can do so only if its velocity is high.

Washington state. This generated a rapidly moving suspension flow that travelled 25 km, and is known to have been rather dilute from witness reports and the way trees were blown down. The blast deposit bears some resemblance to low-aspect-ratio ignimbrites⁷.

Dade and Huppert¹ have now tested in detail the dilute hypothesis in the case of the Taupo ignimbrite. They use the equations for a dilute, fully turbulent, homogeneous suspension current that they have validated using laboratory aqueous flows. They show that the 80-km runout distance of the Taupo flow, as well as radial variations of deposit thickness and grain size, can be explained by a flow with a solid fraction of 0.3 per cent, volume flux $40 \text{ km}^3 \text{ s}^{-1}$ and travel duration 15 minutes. Neither the huge discharge rate nor the flow speed of 200 m s^{-1} are inconsistent with the known properties of pyroclastic flows. But the ability of a dilute, turbulent flow to transport coarse rock debris such large distances and the poor sorting of the theoretical deposit laid down will go against the intuition of many volcanologists.

So the calculations reproduce the essential features of the Taupo ignimbrite, but some uncertainties remain. There are puzzling discontinuities in radial variations of grain size at Taupo, not explicitly addressed by Dade and Huppert, which are qualitatively consistent with a high-concentration flow model⁴. Dense, high-speed solid-gas flows may be highly heterogeneous⁸, however, and cannot yet be modelled, so it is unclear how well they can quantitatively explain the Taupo observations.

Dade and Huppert reinforce their case by showing that clasts of different densities coexisting in the Taupo deposit are in approximate aerodynamic equilibrium with each other, as expected from a low-concentration flow and as observed at Mount St Helens⁷. Ignimbrites characteristically contain both dense rock fragments and frothy, vesicular pumice. During sedimentation from a dilute flow, the largest dense and light clasts deposited at a given site should have the same settling speed in hot gas. In flows of higher concentration, buoyancy forces would permit large pumices to travel fur-

ther than calculated on aerodynamic grounds alone.

Despite its apparent success at Taupo, I think it would be unwise to extrapolate the dilute flow model to ignimbrites of higher aspect ratio. Many show clear evidence of emplacement as dense, fluidized dispersions. For example, at the Valley of Ten Thousand Smokes in Alaska, the pyroclastic flows of 1912 were so concentrated and sluggish that in one place they were unable to surmount a ridge 25 m high⁹. The common occurrence of ignimbrites that are densely welded over thousands of square kilometres¹⁰ implies efficient heat retention during transport that may be hard to reconcile with dilute flow.

GLACIAL CYCLES

Trees retreat and ice advances

Mark Chandler

ONE hundred and fifteen thousand years ago, the Earth began a descent from the warmth of the last interglacial to the frigid climate of the last ice age. What began as a period with air temperatures similar to the present or perhaps slightly warmer, had declined by 21,000 years ago to a state 7–10 °C colder than today. High-latitude temperatures were at least 15 °C colder, and ice sheets more than 1 km thick covered much of North America and large portions of Eurasia above 55° N. The mechanism commonly accepted as initiating the growth of these massive continental ice sheets is the reduction of summertime solar radiation at high latitudes, resulting from cyclical variations in the orbit of the Earth. But many global climate modelling studies have found that the reduction in solar radiation that occurred 115 kyr ago does not, by itself, yield conditions suitable for the maintenance of year-round snow cover. Now, however, climate model experiments reported by Gallimore and Kutzbach on page 503 of this issue¹ show that changing vegetation patterns, specifically the spread of tundra, may have played a prominent role in this reversal of fortunes for the global climate.

In 1989, Rind *et al.*² published an article discussing the inability of a particular 'general circulation model' (GCM) to initiate ice-sheet growth using the conditions of 115 kyr ago. Since then it has been shown that most, if not all, GCMs are similarly limited given solar radiation as the sole forcing change^{3,4}. Speculation that the record is misinterpreted has been widely dismissed given the considerable independent evidence for rapid ice-sheet growth⁵. Concern that these models lack the appropriate sensitivity to simulate even such large climate changes was more seriously considered, because it has impli-

It now appears possible that ignimbrites are generated by a continuum of pyroclastic flows, from small, highly concentrated ones to high-speed types emplaced under dilute, highly turbulent conditions. Quantitative understanding of this spectrum of complex and dangerous two-phase flows presents a major challenge to volcanologists and fluid dynamicists. Dade and Huppert's elegant marriage of mathematics with field observations offers a way forward. □

Tim Druitt is in the Département des Sciences de la Terre (CNRS-URA10), Université Blaise Pascal, 5 Rue Kessler, 63038 Clermont-Ferrand, France.

cations for estimating future climate change.

But the unanimity of negative results for ice initiation bolsters the belief that the models are not fundamentally flawed, but rather are missing pieces of the puzzle. The search for these missing pieces has led climate modellers to incorporate other environmental changes into their experiments. Examples include reducing carbon dioxide levels based on ice-core records and using altered sea surface temperatures with the assumption that ocean circulation was somehow affected. Without exception these added forcings have not yielded permanent snow cover over the regions that eventually came to be dominated by ice sheets.

The new GCM experiments¹ show that tundra expansion near the end of the last interglacial may have affected the Earth's albedo enough to help promote glaciation. In their experiments, Gallimore and Kutzbach use a coarse-resolution version of the NCAR Community Climate Model (CCM1) coupled to a simple mixed-layer ocean. The effects of tundra expansion were examined in two steps, to imitate slowly shrinking high-latitude forests. The expansion of tundra is most important to the radiation balance during winter months, because snow lies in a highly-reflecting layer on the open tundra, whereas dense forests are likely to mask it.

Simulations using reductions in solar radiation and atmospheric carbon dioxide content alone failed to produce perennial snow fields except over the permanent ice caps of Greenland and Antarctica. But the expanded tundra dramatically decreases the snow-free period and increases the area over which snow remains throughout the summer, to accumulate during ensuing winters. Over the Keewatin area and Baffin Island, key glacial initiation

1. Dade, W. B. & Huppert, H. E. *Nature* **381**, 509–512 (1996).
2. Davis, D. K., Quearry, M. W. & Bonis, S. B. *Geol. Soc. Am. Bull.* **89**, 369–384 (1978).
3. Sparks, R. S. J. *Sedimentology* **23**, 147–188 (1976).
4. Wilson, C. J. N. *Phil. Trans. R. Soc. Lond. A* **314**, 229–310 (1985).
5. Wilson, C. J. N. *J. volcan. geotherm. Res.* **8**, 231–249 (1980).
6. Valentine, G. A. *Bull. volcan.* **49**, 616–630 (1987).
7. Druitt, T. H. *Bull. volcan.* **54**, 554–572 (1992).
8. Anilkumar, A. V., Sparks, R. S. J. & Sturtevant, B. *J. volcan. geotherm. Res.* **56**, 145–160 (1993).
9. Hildreth, W. *J. volcan. geotherm. Res.* **18**, 1–56 (1983).
10. Streck, M. J. & Grunder, A. L. *Bull. volcan.* **57**, 151–169 (1995).

centres, snow accumulates at a rate of 30 to 50 cm of water equivalent per year. Over thousands of years this would lower sea level by at least as much as isotopic records indicate.

Although the news of a modelling success is encouraging, the authors point out that their model accumulates large amounts of snow in regions that glaciologists believe were relatively ice free during the last ice age. The only other GCM that reported an ability to initiate ice sheets³ had a similar problem.

Another problem that remains unresolved is how the continental glaciers were able to extend to latitudes south of 60° N. Further north, the extrapolated accumulations would rapidly lead to absurdly thick ice sheets unless the ice began to expand southward across the continents. But in even the most encouraging experiments, the region of perennial snow cover only begins 20 degrees north of the Laurentide glacier's southernmost margin, and the new expanded-tundra experiments have land south of 55° N snow free for over 3 months out of the year. It is difficult to see how the glaciers can have survived so far south. Of course, the ice sheet may create local climatic effects that support continued ice growth. Furthermore, rapid snow accumulation at high latitudes might enable a rapidly flowing glacier to reach the mid-latitudes before melting. But pursuing such hypotheses will require lengthy simulations with a coupled GCM/ice-sheet model.

Whatever the answer, the distinctions between various model results are not nearly as glaring as one might expect from their contrary conclusions. A difference of only one degree in key regions can be the difference between ice-sheet initiation and melting snow. Biases of this magnitude are common in current climate runs and, because the results depend on reaching critical thresholds, disparate conclusions are bound to exist. Given the uncertainties in cloud distributions and sea-ice formation, the sensitivity of the models to comparable forcings is remarkably similar. Unfortunately, as far as ice-sheet growth is concerned, their sensitivity remains inadequate.

Obvious extensions to the latest experiments include using land surface schemes that more realistically, or even interactively, simulate vegetation and ground hydrology interactions. The transformation from forest to tundra is extreme in many respects — not only albedo is altered during such a changeover. Soil

moisture, runoff and surface roughness all generate feedbacks that may affect the growth of glaciers. Ocean thermohaline circulation, other greenhouse gases (methane for example), and desert dust aerosols have all been suggested as important influences on ice-sheet initiation and maintenance. Glacial cycles have even been attributed to external mechanisms, such as periodic variation in the flux of interplanetary dust particles.

The effects of most of these factors re-

main unexamined, and Gallimore and Kutzbach's experiments demonstrate once again that, when it comes to the study of climate change, previously overlooked feedbacks may prove more important than the original instigator of change. □

Mark Chandler is at the Goddard Institute for Space Studies, Columbia University, 2880 Broadway, New York, New York 10025, USA.

NEUROBIOLOGY

Agrin signals at the junction

Clarke R. Slater

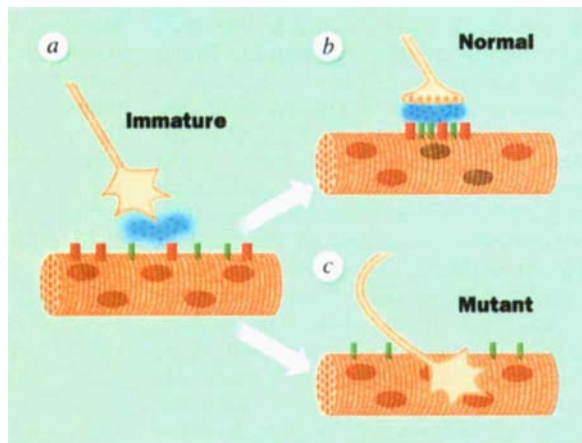
THE ability of the nervous system to deal with the real world depends on rapid communication between its neurons at synaptic junctions. That communication involves the release of a chemical transmitter from one neuron and its binding to a receptor protein on the next, so the speed of transmission depends on the receptors being concentrated close to the sites of transmitter release.

The control of receptor clustering at synapses during development is believed to involve a second, more leisurely, form of signalling between neurons and their targets. Three papers published in *Cell* on 17 May help to define some of the signals involved. Gautam *et al.*¹ have used gene targeting in mice to show that agrin, a neurally derived protein that induces the aggregation of acetylcholine receptors (AChRs) on muscle cells *in vitro*, is also essential for the formation of neuromuscular junctions *in vivo*. DeChiara *et al.*² have used a similar approach to show that MuSK, a muscle-specific protein kinase, is also essential. Connecting these observations, Glass *et al.*³ provide evidence that MuSK mediates the response to agrin.

For practical reasons, control of the distribution of synaptic receptors is best understood at the neuromuscular junction, where the principles of chemical synaptic transmission were first elucidated. AChRs are expressed all over the surface of immature muscle fibres but aggregate at the primitive neuromuscular junction soon after the nerve makes contact. The immature AChR cluster re-

quires the presence of the nerve for a few days to acquire stability. After that, the cluster persists if the nerve is destroyed, implying that information adequate to maintain junctional AChR clusters becomes 'built in' to the surface of the muscle fibre during synapse formation.

During the past decade, McMahan and his colleagues have obtained striking support for this view (see ref. 4 for review). They first found that molecules able to cause aggregation of AChRs on cultured muscle cells are bound to the region of the basal lamina, a loosely knit sheath of extracellular material that lies between nerve and muscle. They then identified a family of proteins, some of which account



Proposal for synapse formation at the neuromuscular junction. *a*, In early development, before the motor axon of the nerve has reached the muscle fibre, the fibre lacks local specialization. Muscle-specific protein kinase (MuSK, red), acetylcholine receptors (AChR, green) and other proteins are uniformly distributed along the fibre. The motor axon releases agrin (blue), which activates MuSK at the site of contact. *b*, In later development, agrin accumulates in the basal lamina of the synapse. MuSK, AChRs and many other molecules (not shown) become concentrated in the mature synapse, due in part to local transcription from synaptic nuclei (shaded). The motor axon differentiates a presynaptic terminal in response to unidentified signals from the muscle fibre. *c*, In mutants lacking agrin or MuSK, the motor axon fails to induce postsynaptic specialization in the muscle fibre; the fibre does not produce the stop/differentiate signal, so motor axons wander over the muscle surface without differentiating.

- Gallimore, R. G. & Kutzbach, J. E. *Nature* **381**, 503–505 (1996).
- Rind, D., Peteet, D. & Kukla, G. *J. geophys. Res.* **94**, 12851–12871 (1989).
- Dong, B. & Valdes, P. J. *J. Clim.* **8**, 2471–2496 (1995).
- Phillips, P. J. & Held, I. M. *J. Clim.* **7**, 767–782 (1994).
- Peteet, D., Rind, D. & Kukla, G. *Geol. Soc. Am. Spec. Pap.* **270**, 53–69 (1992).

for this activity, and named it agrin. Agrins are made by motor neurons and transported to the periphery where they become highly concentrated in the basal lamina at the neuromuscular junction. Crucially, for the hypothesis that agrin mediates nerve-induced AChR aggregation⁴, they showed that the formation of AChR clusters during neuromuscular junction formation *in vitro* is prevented by agrin-binding antibodies⁵.

One way of testing the importance of agrin *in vivo* is to knock out the gene concerned. Gautam *et al.*¹ have now done this, with striking results. Mice lacking agrin are immobile when born and die within hours. Although AChRs are present on the surface of the muscle fibre, they are not aggregated into clusters at sites of nerve-muscle contact. Although motor nerve fibres branch exuberantly within the muscles, they do not form recognizable neuromuscular junctions. Furthermore, the muscle cell nuclei underlying the nerve branches, which normally preferentially express the genes that encode the AChR, fail to do so in the absence of agrin. These studies establish the essential role of agrin in synapse formation and add support to the view that its effects go beyond those on AChR aggregation⁶.

If agrin is an essential signal for neuromuscular junction formation, what mediates the muscle's response to it? Considerable excitement accompanied the finding that α -dystroglycan, part of a complex that links dystrophin to the cell surface, is a major agrin-binding protein. But it has since been shown that the region of the agrin molecule by which it binds to α -dystroglycan is not essential for AChR aggregating activity and that isoforms of agrin made by muscle cells bind well to α -dystroglycan but do not cause AChR aggregation⁷. So although α -dystroglycan binds agrin, it seems that it is not the receptor that mediates AChR clustering.

It has been known for some time that agrin activity requires protein phosphorylation⁸, suggesting that agrin activates a protein kinase. Last year, Valenzuela *et al.*⁹ identified a receptor tyrosine kinase specific for skeletal muscle, which they named MuSK and which has some particularly intriguing properties. Like AChRs, MuSK is uniformly distributed over the surface of immature and inactive muscle fibres, and becomes restricted to the region of the neuromuscular junction during development. Now the same group² reports that mice in which the MuSK gene has been knocked out have a phenotype that is virtually identical to that of agrin-deficient mice and a strikingly similar and equally comprehensive impairment of the development of neuromuscular junctions. Testing the notion that MuSK might itself be the agrin receptor, Glass *et al.*³ found that AChRs on MuSK-deficient muscle

cells grown in culture did not aggregate in response to agrin.

Receptor tyrosine kinases respond to the binding of extracellular signalling molecules such as growth factors by phosphorylating proteins, which may include themselves. Fragments of agrin that are adequate to cause AChR aggregation also cause rapid autophosphorylation of MuSK, whereas a number of other growth factors do not³. However, agrin does not bind directly to the extracellular domain of MuSK on its own, either *in vitro* or when MuSK expression is induced in non-muscle cells. On the other hand, expression of MuSK in immature myotubes induces agrin binding to the cell surface. This suggests that, as with a number of other receptor tyrosine kinases, agrin binds to a complex of MuSK with at least one other component which is present specifically on muscle cells.

It has been a long haul to these genetic knockout studies, and it is highly satisfying that the results are now in: they provide strong support for McMahan's 'agrin hypothesis'⁴, and do much to establish that agrin is an essential signal for AChR aggregation at the neuromuscular junction. Although the evidence that MuSK is the receptor that mediates these effects of agrin is not yet unequivocal, it is clear that it plays an essential part in the muscle's response to agrin. Thus signal-activated protein phosphorylation, a widespread component of growth control systems, also plays a key part in synapse formation.

In addition, these papers and another in press¹⁰ imply that the consequences of agrin signalling extend beyond AChR aggregation to other aspects of the development of neuromuscular junctions, including the control of nerve growth and AChR gene expression. Because agrin and/or its gene is expressed in both cholinergic and non-cholinergic neurons in the central nervous system¹¹, it is likely that there, too, agrin provides essential signals for synapse formation. □

Clarke R. Slater, of the Department of Neurobiology, University of Newcastle upon Tyne, is in the Muscular Dystrophy Group Research Laboratories, Newcastle General Hospital, Westgate Road, Newcastle upon Tyne NE4 6BE, UK.

- Gautam, M. *et al.* *Cell* **85**, 525-535 (1996).
- DeChiara, T. M. *et al.* *Cell* **85**, 501-512 (1996).
- Glass, D. J. *et al.* *Cell* **85**, 513-523 (1996).
- McMahan, U. J. *Cold Spring Harb. Symp. quant. Biol.* **60**, 407-418 (1990).
- Reist, N. E., Werle, M. J. & McMahan, U. J. *Neuron* **8**, 865-868 (1992).
- Cohen, I. *et al.* *Soc. Neurosci. Abstr.* **21**, 1748 (1995).
- Meier, T., Gesemann, M., Cavalli, V., Ruegg, M. A. & Wallace, B. G. *EMBO J.* (in the press).
- Wallace, B. G. *J. Cell Biol.* **125**, 661-668 (1994).
- Valenzuela, D. M. *et al.* *Neuron* **15**, 573-584 (1995).
- Jones, G. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- O'Connor, L. T., Lauterborn, J. C., Gall, C. M. & Smith, M. A. *J. Neurosci.* **14**, 1141-1152 (1994).

Electric fire

ALONE among vehicles, the rocket has to carry its entire energy supply — oxygen as well as fuel. In this connection Daedalus recalls the wire-guided missile, steered by commands sent to it along an unreeling cable. In principle, it could get energy from the cable as well. But how to power a rocket electrically? The plasma torch attains a temperature far beyond that of combustion by means of an electric arc struck across its flame. Daedalus reckons that a rocket might similarly double the energy of its fuel by striking a big plasma arc across its exhaust jet. Sadly, the required heavy current would need such thick unreeling cables that they would soon drag the vehicle back to Earth.

But Daedalus does not give up so easily. The plasma torch is hot enough to ionize many of its gas molecules. So he is designing a rocket whose exhaust plume is itself an electrical conductor. For maximum conductivity, he will dope the fuel with a readily ionized element such as caesium. The vehicle will need two separate and widely spaced motors, one for the outward current and one for the return, but this is a minor complication. There will be two flat electrodes on the launch pad, one under each motor, and both connected to a power station. When the motors fire, the circuit will be completed, and a huge current will pass up one plume and down the other. The rocket will roar into the sky, trailing two conducting plumes that carry energy to its motors. The current will heat the whole length of both plumes, keeping them hot and conducting, but wasting a lot of energy in the process. Even so, the heat will be most intense at the point of minimum cross-section — in the rocket nozzles themselves, just where it is needed. In any case, electricity is very cheap compared to rocket fuels. The system could be quite economic.

How high could the rocket climb before its energizing current chokes off? A fluid conductor is compressed and compacted by the magnetic 'pinch effect'. Daedalus argues that this will oppose the diffusive spreading of the plumes, and will keep them compact, hot and conducting for many kilometres into the sky. (On the same principle the two plumes, carrying current in opposite directions, will repel each other, thus preventing a short-circuit between them.) A rocket burns most of its fuel in the early part of its climb; by the time its current is cut off, it will have gained a powerful boost. Furthermore, thrust adjustments and control signals for the tricky early stages of flight could be sent up the conducting plumes. Space launchers will take a new great leap upwards.

David Jones

Neolithic resinated wine

SIR — We report here that we have found the earliest chemical evidence of wine. We have analysed a pottery jar from a Neolithic village in Iran's northern Zagros mountains which contained wine, with a *Pistacia* tree resin additive. The jar was produced in about 5400–5000 BC, two thousand years earlier than previous evidence from the earliest civilization of the Near East¹. Our new evidence belongs to the period when the first permanent human settlements, based on domesticated plants and animals as well as minor crafts such as pottery-making, were being established. It has important implications for the origins of viticulture and vinicul-

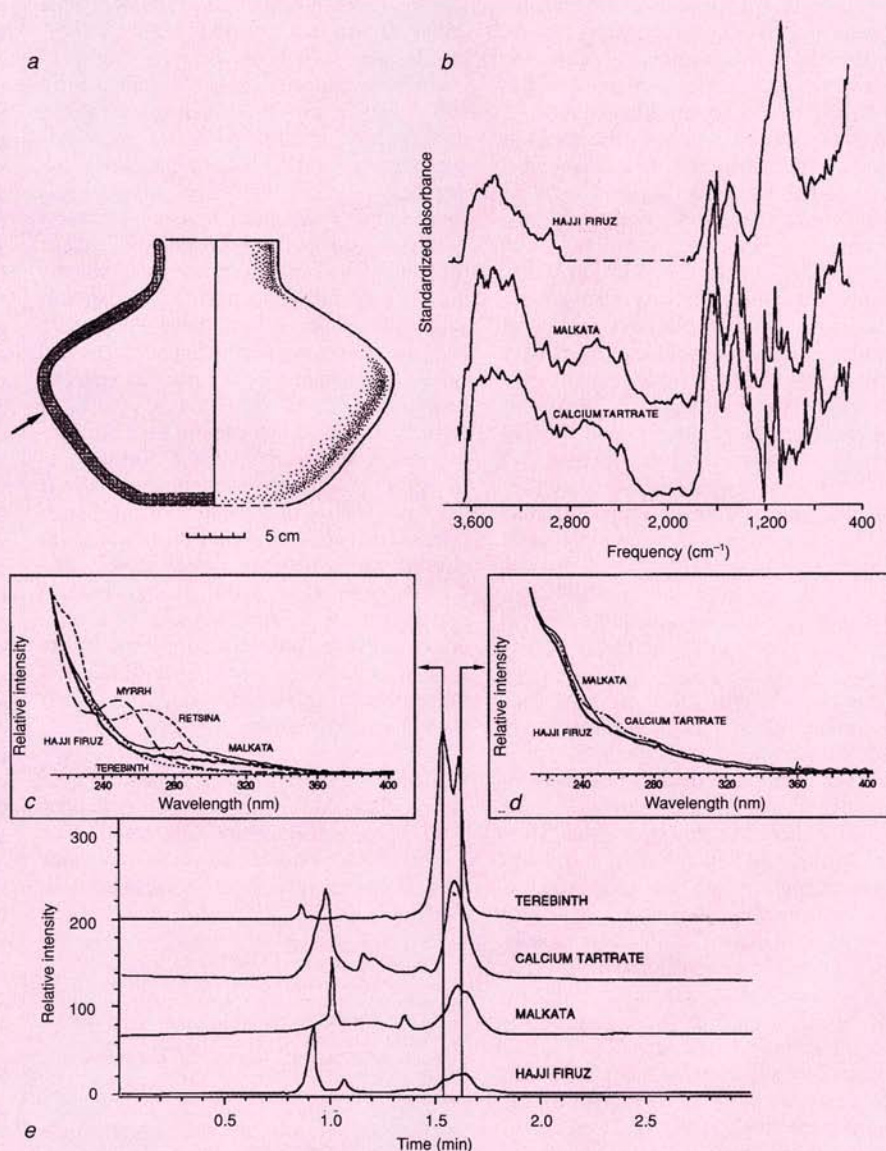
ture, as well as for the development of modern diet, medical practice and society generally².

One of us (M.M.V.) excavated the Neolithic jar (8.81; *a* in the figure), which has an interior yellowish residue, in the 'kitchen' of a square mud-brick building at the site of Hajji Firuz Tepe³. The jar mainly contained the calcium salt of tartaric acid, as established by infrared spectrometry (*b* in the figure), liquid chromatography (*c–e*) and a wet-chemical test⁴. Also identified is the yellowish oleoresin of the *Pistacia atlantica* Desf. terebinth tree (*e*), which grows abundantly throughout the Near East⁵ and was widely used

as a medicine and wine additive in antiquity⁶. The presence of combined tartaric acid/tartrate and terebinth resin has been confirmed for a large group of ancient Near Eastern jars, including Egyptian amphoras with wine dockets⁷.

Tartaric acid occurs naturally in large amounts only in grapes, and was converted to the insoluble calcium salt in the calcareous environment of the site. We think that the jar originally contained a liquid, because it has a long, narrow neck and the interior residue is confined to its bottom half. Under normal conditions and at room temperature, grape juice quickly ferments to wine⁸. Because of slow pressing methods in antiquity and high temperatures in the Middle East, fermentation had probably begun before the jar was

Diffuse-reflectance Fourier-transform infrared (IR) spectra (*b*) and high-performance liquid chromatograms (*e*) and ultraviolet (UV) spectra (*c, d*) of methanol extracts of the Hajji Firuz Tepe jar (*a*, arrow marks sampling area) and ancient and modern reference samples. In *b*, calcium tartrate accounts for the IR maxima at 1,613 and 1,561 cm^{-1} , due to the carboxylate group, in the Hajji Firuz sample. Many other absorption bands (for example, 3,596; 3,523; 3,442; 1,385; 1,330; 1,275; 713; 596; 555; and 484 cm^{-1}) match and have similar shapes and intensities. The broad absorption centred at about 3,450–3,400 cm^{-1} results from hydroxyl groups and water of hydration, whereas the inorganic clay band (maximum at 1,032 cm^{-1}) masks characteristic peaks between 1,250 and 800 cm^{-1} , which are visible in an ancient reference sample, a fourteenth century BC wine amphora from the palace of Pharaoh Amenhotep III in Thebes, Egypt¹¹, at 1,060, 1,005, 955 and 815 cm^{-1} . The same three samples show comparable chromatographs (*e*) and UV spectra (*d*) at a retention time of 1.62 min. The presence of *Pistacia atlantica* Desf. terebinth resin, composed of characteristic triterpenoids¹², in the Hajji Firuz sample is supported by the hydrocarbon peaks at 2,926 and 2,858 cm^{-1} in the IR spectrum (*b*), but more definitively, by the latter's chromatogram (*e*) and those of the ancient and modern reference samples, together with their corresponding UV spectra (*c*) at a retention time of 1.54 min. The *P. atlantica* resin tested was in the form of nodules inside an amphora¹³ of the fourteenth century BC merchantman shipwreck at Uluburun. Modern myrrh, also a wine additive in antiquity, and Greek retsina wine, in which sandarac (*Tetraclinis articulata*) resin is the primary additive, show different UV absorptions in the 230–290-nm range (*c*) from that of *P. atlantica*. For sherd extraction and IR analytical procedures, see refs 1, 2. Because of its small sample size (4 mg), the Hajji Firuz IR spectrum was obtained by a micro technique. The liquid chromatograph was run at ambient temperature, using a 25 cm $\times$ 4.6-mm silica column, a flow rate of 2.0 ml per min, and an UV detector over the 200–400-nm range. Variation in the retention times of earlier peaks (0.9–1.1 min) on some chromatograms (*e*) is probably due to the different temperatures of individual analyses, as well as interactions in complex mixtures.



filled⁹. The addition of terebinth resin, which is readily soluble in alcohol, served in part to disturb and inhibit the growth of bacteria (*Acetobacter*) that convert wine to vinegar, besides masking any offensive taste or odour⁶. The jar's dense fabric and stoppering further assured the preservation of the wine.

Hajji Firuz Tepe lies within the ancient and modern distributional zone of the wild grape (*Vitis vinifera* L. subsp. *sylvestris*) and *Pistacia* tree varieties, as established by pollen cores from Lake Urmia¹⁰. Although wild grapes were available seasonally, the wine in the jar might well have been produced from a precursor of the highly successful domesticated type (subsp. *vinifera*), still used to make most modern wine. Many centuries of experimentation, beginning in the Neolithic period, were required to develop a hermaphroditic, more productive grape vine, and to refine the methods of winemaking and storage.

Modern diets include many foods and beverages used in Neolithic times¹¹ which have rarely been identified chemically. Wine was a highly desirable grape product, because of its unique dietary and medical benefits and psychotropic effects. The identification of resinated Neolithic wine is particularly significant, because of this beverage's impact on social customs, religions and economies throughout the world².

Patrick E. McGovern

Donald L. Glusker

Lawrence J. Exner

Museum Applied Science Center for Archaeology (MASCA),

University of Pennsylvania

Museum, Philadelphia,

Pennsylvania 19104-6324, USA

Mary M. Voigt

Department of Anthropology,

College of William and Mary,

PO Box 8795, Williamsburg,

Virginia 23187-8795, USA

1. Michel, R. H., McGovern, P. E. & Badler, V. R. *Analyt. Chem.* **65**, 408A-413A (1993); *Nature* **360**, 24 (1992).
2. McGovern, P. E., Fleming, S. J. & Katz, S. H. (eds) *The Origins and Ancient History of Wine* (Gordon & Breach, Luxembourg, 1995).
3. Voigt, M. M. *Hajji Firuz Tepe, Iran: The Neolithic Settlement* 37-42, 128, 132, 184-185, 348-350, Figs 30 & 87d (field no. 879), cf. pl. 28e (Univ. Museum, Philadelphia, 1983).
4. Feigl, F. *Spot Tests in Organic Applications* 384-385 (Elsevier, New York, 1954).
5. Zohary, M. *Palest. J. Bot., Jerusalem Ser.* **5**, 187-228 (1952).
6. Majno, G. *The Healing Hand: Man and Wound in the Ancient World* 64, 124, 210-225 (Harvard Univ. Press, Cambridge, 1975).
7. Hayes, W. C. *J. Near E. Stud.* **10**, inscr. 41, 35-41, 89, Fig. 6 (1951).
8. Farkaš, J. *Technology and Biochemistry of Wine* (Gordon & Breach, New York, 1988).
9. Lucas, A. & Harris, J. R. *Ancient Egyptian Materials and Industries* 16-18; 92-94; 316-324 (Arnold, London, 1962).
10. Bottema, S. *Rev. Palaeobot. Palynol.* **47**, 241-261 (1986).
11. Katz, S. H. & Voigt, M. M. *Expedition* **28**, 23-34 (1986).
12. Mills, J. S. & White, R. *Archaeometry* **31**, 37-44 (1989).
13. Pulak, C. *Am. J. Archaeol.* **92**, 10-11 (1988).

When batterer becomes murderer

SIR — I reported in Scientific Correspondence an estimate of the probability x that a batterer was the murderer of his mate when she is known to have been murdered by somebody¹. The probability x must be an increasing function of the degree of battery. If one defines a representative batterer as someone who averages one assault per year, one can estimate x relative to such a batterer. Threats of death and other aggressive behaviour are tantamount to assaults, but let us leave something to the discretion of the jury! This communication is not concerned specifically with the O. J. Simpson case, where there are mountains and valleys of further evidence.

Unfortunately, my earlier estimate¹ of x was much too low because it was based on an incorrect statement made by the distinguished lawyer Alan Dershowitz in a television interview in March 1995. (He repeated the statement in another interview recorded on 2 April 1996.) He said that less than a tenth of 1 per cent of batterers go on to murder their wives. He did not say at the time that this fraction is really the batterer's probability of murder per assault. For a representative batterer it is approximately the probability of murder per year, which is much less than the probability of his committing murder sooner or later. I have found that Dershowitz² had made the same mistake, but his raw data are presumably reliable³.

On page 101 of his recent bestselling book⁴, Dershowitz writes, "We knew we could prove, if we had to, that an infinitesimal percentage — certainly fewer than 1 out of 2,500 — of men who slap or beat their domestic partners go on to murder them." He used this and similar estimates to argue that battery should be regarded as inadmissible evidence. But on page 104 he mentions that, in 1992, there were 1,432 murders among 2-4 million assaults. So our representative batterer's probability of murder is about 1/2,000 per year. His probability of going on to murder is an order of magnitude larger, hardly "infinitesimal". Even if Dershowitz's estimate had been right, one could equally argue that the knowledge that an accused was the husband of the victim should be inadmissible because husbands very seldom murder their wives!

Let G mean that the accused is factually guilty (committed the crime in, say,

1994). Let M mean that the woman was murdered by somebody in 1994. Let Bat mean that she had been battered by her mate about once per year for several years. Let the given information be M and Bat . We would like to estimate $x = P(G|M \text{ and } Bat)$, where the vertical stroke is the standard notation for 'given' or 'conditional upon'.

Our conclusion will be that, for a standardized batterer, x is approximately 0.9 instead of the 0.5 I estimated earlier. Both my earlier contribution and this one use the odds form of Bayes's theorem⁵⁻⁸. The odds corresponding to a probability P can be defined as $P/(1-P)$. The theorem gives

$$\frac{O(G|M \text{ and } Bat)}{O(G|Bat)} = \frac{P(M|Bat \text{ and } G)}{P(M|Bat \text{ and not } G)} \quad (1)$$

The left side is the Bayes factor (in favour of guilt provided by the evidence M , given Bat). Its numerator is x , and its denominator is about 1/2,000 because the year of the murder is specified. The numerator of the right side is 1, and we now estimate its denominator.

There are about 25,000 murders per year in the US population of about 250,000,000 (p. 964 of ref. 9), giving a probability of 1/10,000 of being murdered in a given year. But approximately a quarter instead of a half of murdered victims are female (p. 283 of ref. 3), so being female reduces this probability to 1/20,000. Thus:

$$\begin{aligned} P(M|Bat \text{ and not } G) &= \\ P(M|\text{not } G) &\approx 1/20,000 \end{aligned} \quad (2)$$

Therefore, the right-hand side of equation (1) is about 20,000. This, then, is the Bayes factor in favour of true guilt and

$$\begin{aligned} O(G|M \text{ and } Bat) &\approx 20,000/2,000 \\ &= 10 \end{aligned} \quad (3)$$

so the corresponding probability is approximately 0.9 as I claimed.

(Some refinements and more details are given in an unpublished technical report, available on request.)

I. J. Good

Department of Statistics,

Virginia Polytechnic Institute and

State University,

Blacksburg, Virginia 24061-0439, USA

1. Good, I. J. *Nature* **375**, 541 (1995).
2. Dershowitz, A. M. *Los Angeles Times* Sect. M, p. 5 (15 January 1995).
3. *Uniform Crime Reports for the United States and its Possessions* (Int. Ass. Chiefs of Police, Washington DC, 1993).
4. Dershowitz, A. M. *Reasonable Doubts: The O.J. Simpson Case and the Criminal Justice System* (Simon & Schuster, New York, 1996).
5. Wrinch, D. & Jeffreys, H. *Phil. Mag. Ser. 6*, 369-390 (1921).
6. Good, I. J. *Probability and the Weighing of Evidence* Ch. 6 (Griffin, London, 1950).
7. Good, I. J. in *The Use of Statistics in Forensic Science* (eds Aitken, C. G. G. & Stoney, D. A.) Ch. 3 (Horwood, Chichester, 1991).
8. Good, I. J. *Good Thinking* (Univ. Minnesota Press, 1983).
9. Farmigetti, R. (ed.) *World Almanac* (Funk & Wagnalls, Mahwah, NJ, 1994).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and five references.

Plastid in human parasites

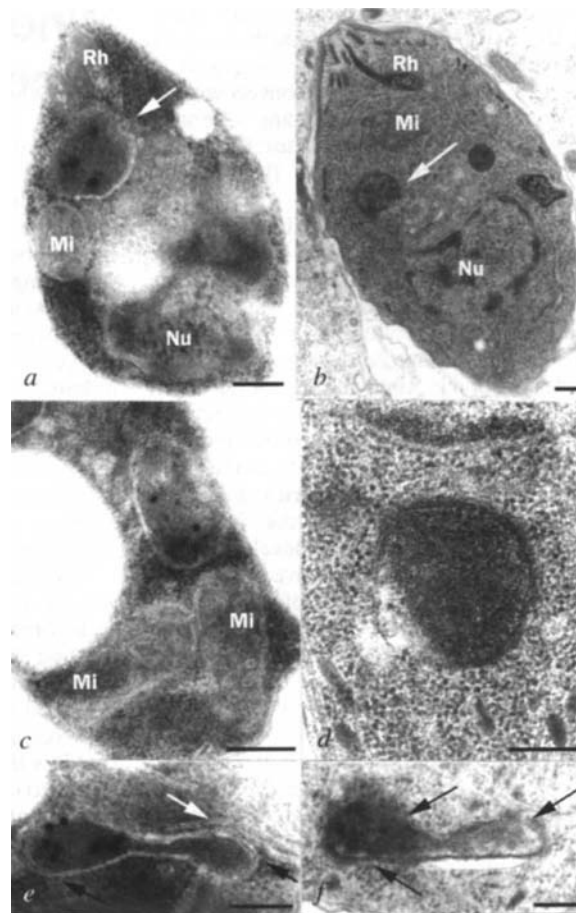
SIR — The discovery in malarial and toxoplasmodial parasites of genes normally occurring in the photosynthetic organelle of plants and algae has prompted speculation that these protozoans might harbour a vestigial plastid¹. The plastid-like parasite genes occur on an extrachromosomal, maternally inherited², 35-kilobase DNA circle with an architecture reminiscent of that of plastid genomes^{3,4}. Although the 35-kb genome is distinct from the 6–7-kb linear mitochondrial genome^{3–6}, it is not known where in the parasite cells the plastid-like genome resides.

To determine whether a plastid is present, we used high-resolution *in situ* hybridization⁷ to localize transcripts of a plastid-like 16S ribosomal RNA gene from *Toxoplasma gondii*⁸, the causative agent of toxoplasmosis. Transcripts accumulate in a small, ovoid organelle located anterior to the nucleus in the mid-region of the cell (*a, b* in the figure). We observe only one of these organelles per cell (*a–c*), except during division (endodyogeny), when an organelle is associated with each forming daughter cell (not shown). At the onset of endodyogeny, the organelle assumes a cylindrical shape (not shown) and eventually acquires a constriction, creating a dumbbell shape (*e, f* in the figure) which presumably divides into two daughter organelles. The organelle is distinct from the tubulocristate mitochondria (*c*). At least two membranes (some profiles possibly indicate a third membrane) surround the organelle (*b–f*). Certain images (*e* in the figure) reveal a layer of endoplasmic reticulum (sometimes completely) enveloping the organelle, creating a misleading impression of multiple (typically four or five) surrounding membranes.

Particles of 18 nm diameter (approximately the size of bacterium-like 70S ribosomes) are the only discernible structures within the organelle (*d, f*), the contents

being otherwise homogeneous. The 35-kb genome-containing organelle identified here did not escape the attention of early electron microscopists who — not expecting the presence of a plastid in a protozoan parasite like *Toxoplasma* — ascribed to it various names, including 'Hohlzylinder' (hollow cylinder), 'Golgi adjunct' and 'große Vakuole mit kräftiger Wandung' (large vacuole with stout surrounds) (see refs cited in ref. 9). Our preliminary experiments with *Plasmodium falciparum*, the causative agent of the most lethal form of malaria, identify an organelle (not shown) which appears similar to the *T. gondii* plastid. The number of surrounding membranes in the *P. falciparum* plastid, and its relationship to the previously described spherical body of *Plasmodium* (see ref. 9), are unknown. Our *in situ* hybridization data thus identify a third, hitherto putative, genetic compartment (a plastid) in these parasite cells. Because the plastid genome encodes components of ribosomes and the translation process^{4,10}, and we observe ribosome-like particles in it, the plastid compartment probably contains machinery to express its information content.

Toxoplasma and *Plasmodium* belong to phylum Apicomplexa, which is closely related to dinoflagellate algae (refs in ref. 3), so the parasite plastid could be a non-photosynthetic derivative of the dinoflagellate plastid (but see ref. 5). Unfortunately, no genes have yet been characterized from dinoflagellate plastids, so molecular comparison of dinoflagellate and apicomplexan plastids must wait. Whatever its origin, the presence of a plastid in apicomplexan parasites probably explains their sensitivity to certain herbicides (ref. 11 and refs therein) and drugs inhibiting plastid metabolism^{12,13}. The plastid is thus a welcome new, parasite-specific target for therapeutic agents. The role of the plastid in obligate intracellular parasites is completely unknown, but our identification of the organelle is a first step towards isolation for biochemical analysis to



a, Localization of transcripts from a plastid-like 16S rRNA gene⁸ in tachyzoite (longitudinal section) of *Toxoplasma gondii* RH by high-resolution *in situ* hybridization. Colloidal gold markers show transcripts accumulated in an ovoid compartment (arrow). Nu, nucleus; Rh, rhoptries; and Mi, mitochondria. *b*, Standard electron micrograph (embedding courtesy of D. Lindsay, Auburn University) similar to *a*, showing the plastid (arrow). *c*, Mitochondria (Mi) with tubular cristae are not labelled. *d*, Close view of plastid showing at least two surrounding membranes, homogeneous contents and ribosome-like particles (about 18 nm diameter). *e*, Dumbbell-shaped (possibly dividing), double-membrane-bound (black arrows) *Toxoplasma* plastid. Endoplasmic reticulum appressed to the plastid (white arrow) creates impression of extra surrounding membranes. *f*, Standard electron micrograph showing dumbbell-shaped *Toxoplasma* plastid with two surrounding membranes (black arrows) and ribosome-like particles (about 18 nm diameter) within. Scale bars, 0.2 μ m.

complement molecular-genetic approaches already under way^{4,10}.

Geoffrey I. McFadden

*Plant Cell Biology Research Centre,
School of Botany,
University of Melbourne,
Victoria 3052, Australia*

Michael E. Reith

*Janet Munholland
Institute for Marine Biosciences,
National Research Council,
Halifax B3H 3Z1,
Nova Scotia, Canada*

Naomi Lang-Unnasch

*Division of Geographic Medicine,
Department of Medicine,
University of Alabama at Birmingham,
Birmingham, Alabama 35294-2170, USA*

1. Wilson, R. J. M. *Nature* **366**, 638 (1993).
2. Creasey, A. et al. *Molec. biochem. Parasit.* **65**, 95–98 (1994).
3. Palmer, J. D. *Curr. Biol.* **2**, 318–320 (1992).
4. Preiser, P., Williamson, D. H. & Wilson, R. J. M. *Nucleic Acids Res.* **23**, 4329–4336 (1995).
5. Wilson, R. J. M., Williamson, D. H. & Preiser, P. *Infect. Ag. Dis.* **3**, 29–37 (1994).
6. Feagin, J. E. A. *Rev. Microbiol.* **48**, 81–104 (1994).
7. McFadden, G. I. in *Electron Microscopy of Plant Cells* (eds Hall, J. L. & Hawes, C. R.) 219–255 (Cambridge Univ. Press, 1991).
8. Egea, N. & Lang-Unnasch, N. *J. eukaryot. Microbiol.* **42**, 679–684 (1995).
9. Siddall, M. E. *Parasit. Today* **8**, 90–91 (1992).
10. Beckers, C. J. M. et al. *J. clin. Invest.* **95**, 367–376 (1995).
11. Hackstein, J. H. P. et al. *Parasit. Res.* **81**, 207–216 (1995).
12. Strath, M. et al. *Trans R. Soc. Trop. med. Hyg.* **87**, 211–216 (1993).
13. Pukrittayakamee, S. et al. *Antimicrob. Ag. Chemother.* **38**, 511–514 (1994).

The birth of microsatellites

SIR — Microsatellites, very short (2–5 base-pair) tandem nucleotide repeats, are found scattered throughout the genomes of eukaryotes^{1,2}. Those loci with a high number of repeat units tend to vary in repeat number among individuals in a species, making them useful in many types of modern genetic studies³. Microsatellites have gained notoriety through their association with certain genetic diseases in humans^{4,6}. Much has been written about the probable mechanisms causing length variation within species^{2,6,7}, but little is known about the evolutionary origins of microsatellites¹. Here we document the independent 'birth' of two separate microsatellites in a short region of the η -globin pseudogene in primates.

As shown in the figure, in many species of primates the DNA sequence in this region is a single nucleotide substitution away from the creation of two different types of simple repeat sequences. Within the hominoids, a single nucleotide substitution of a G to an A occurred along the lineage leading to the common ancestor of the African apes (gorilla, bonobo and chimpanzee) and humans. This substitution created a tetranucleotide repeat, (ATGT)₂, which is expanded to (ATGT)₄ in the African apes and to (ATGT)₅ in humans. As only one sequence is available from each species, we do not know if intraspecific variation in repeat number exists at this locus. However, such polymorphism is unlikely, because loci with fewer than about 5 to 8 repeat units are almost never polymorphic in humans^{1,8}.

A separate nucleotide substitution occurred within the New World monkeys along the lineage leading to owl monkey. This A-to-G transition created a dinucleotide repeat, (GT)₅ or (TG)₆, which has expanded by one repeat unit in this species. η -Globins from other New World monkeys that are more closely related to this owl monkey need to be sequenced to trace the evolution of this new microsatellite more completely, especially to look for additional expansions.

Even though a wide phylogenetic spread of primates has three contiguous GTs and four contiguous TGs in this region of the η -globin pseudogene, of the species sampled, only the owl monkey has a dinucleotide expansion. Thus, a minimum number of repeat units may be necessary before initial expansion occurs. In this dinucleotide repeat case, the minimum number appears to be four or five, and in the tetranucleotide repeat case, two. The evolutionary coupling of the nucleotide substitutions and the expansions suggests that the point mutations created enough repeat units for subsequent microsatellite expansion. The expansions we have found in this region of the η -globin gene are consistent with a stepwise mutation model for microsatellite evolution^{1,2} in which short arrays increase or decrease in repeat number by one or two repeats, probably by replication slippage. We propose that initial expansion of short repeat sequences, such as we document here, occurs relatively slowly on an evolutionary timescale.

Once a critical number of repeat units^{1,8} has arisen in a given species, that locus can become hypervariable, with mutations occurring even on a generational timescale⁷. A broader understanding of microsatellite birth and early evolution will require the sequencing of regions surrounding other expanded loci and analysis of these data in proper phylogenetic context.

Walter Messier

Shou-Hsien Li

Caro-Beth Stewart

*Department of Biological Sciences,
University at Albany,
State University of New York,
Albany, New York 12222, USA*

- Valdes, A. M., Slatkin, M. & Freimer, N. B. *Genetics* **133**, 737–748 (1993).
- Charlesworth, B. *et al. Nature* **371**, 215–220 (1994).
- Dowling, T. E. *et al. in Molecular Systematics* (eds Hillis, D. M. *et al.*) Ch. 8 (Sinauer, Sunderland, MA, 1996).
- Burke, J. R. *et al. Nature Med.* **2**, 347–350 (1996).
- Campuzano, V. *et al. Science* **271**, 1423–1427 (1996).
- Kuhl, D. P. A. & Caskey, C. T. *Curr. Opin. Genet. Dev.* **3**, 404–407 (1993).
- Rubinshtein, D. C. *et al. Nature Genet.* **10**, 337–343 (1995).
- Armour, J. A. L. *et al. Hum. molec. Genet.* **3**, 599–605 (1994).
- Koop, B. F. *et al. Nature* **319**, 234–238 (1986).

Cosmological constant

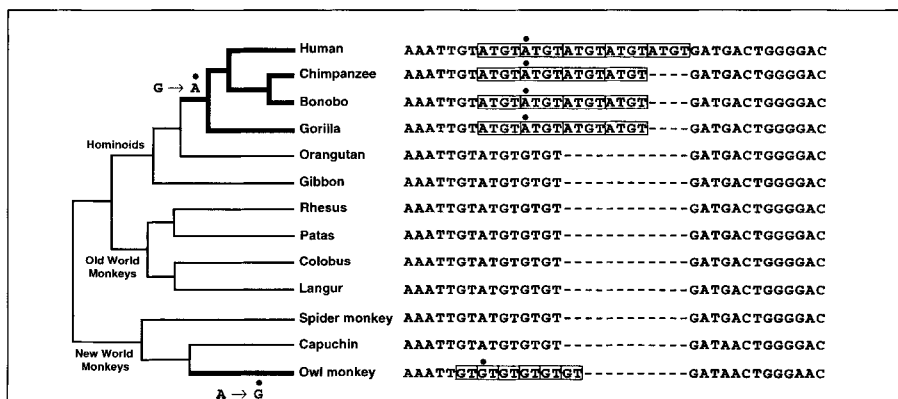
SIR — Taken in sum¹, an impressive variety of observations indicate that the cosmological constant (Λ) is not zero. Principal among these is the existence of stars that are more than 15 billion years old², in a Universe which is apparently much younger (less than 12 billion years old), based on recent determinations of the Hubble constant³. A cosmological constant increases the age of the Universe, but only modestly in a flat Universe. Even with a cosmological constant of $\Omega_\Lambda = 0.7$ (where Ω_Λ is expressed as a fraction of the critical density), concordance between these ages relies heavily on measurement errors^{2,3}.

Moreover, the cosmological constant has a strong effect⁴ on the deceleration parameter q_0 , which provides an independent test of concordance. In particular, $q_0 = 1/2\Omega_m - \Omega_\Lambda$, where Ω_m is the density of matter as a fraction of the critical density. If $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ as suggested¹, then $q_0 = -0.55$. Although presently fraught with uncertainty, attempts⁵ to measure q_0 directly usually suggest positive, rather than negative, values. If $q_0 \geq 0$ in a flat Universe, $\Omega_\Lambda \leq 1/2\Omega_m$, and the cosmological constant cannot be significant.

Stacy S. McGaugh

*Department of Terrestrial Magnetism,
Carnegie Institution of Washington,
Washington, DC 20015, USA*

- Ostriker, J. P. & Steinhardt, P. J. *Nature* **377**, 600–602 (1995).
- Boite, M. & Hogan, C. J. *Nature* **376**, 399–402 (1995).
- Freedman, W. L. *et al. Nature* **371**, 757–762 (1994).
- Peebles, P. E. J. *Principles of Physical Cosmology* (Princeton Univ. Press, 1993).
- Sandage, A. R. in *The Deep Universe*, Saas-Fee Advanced Course 23 (Springer, New York, 1995).



A short portion of a longer alignment of all available nucleotide sequences of the η -globin pseudogene of primates. The multiple alignment of these primate sequences is unambiguous surrounding this region (which encompasses nucleotides 1,311–1,350 of the full-length human sequence, and would be in the second intron of an expressed globin gene). Gaps relative to the human sequence are indicated by dashes. Those nucleotide substitutions (indicated by dots) that created repeat units are mapped onto the known phylogeny of these primates in the manner that requires the fewest events. Boxes delineate microsatellite repeats. For the sake of clean alignment to the human sequence, we have boxed the GT repeats in the owl monkey; however, it is just as likely that this expansion involved a TG repeat, as mutations in the hominoids should have no effect on those in the New World Monkeys. GenBank accession numbers: Human, X02133; chimpanzee, K02542; bonobo, X79724; gorilla, K02543; orangutan, M18038; gibbon, M54985; owl monkey, X02142; spider monkey, J03050; white-fronted capuchin, M81409. Rhesus macaque sequence is from ref. 9. Sequences from patas monkey, Angolan colobus and Hanuman langur were generated in our laboratory by P. Banerjee; analysis of these pseudogenes will be published elsewhere.

Ancient DNA and island endemics

SIR — Most recently extinct and currently endangered species of birds are inhabitants of islands, where the effects of anthropogenic predation and habitat modification have been most severe¹. Recently, palaeontological records have shown that the effect of prehistoric humans on insular biotas, particularly in the Pacific, were far more severe than previously believed^{2–4}, and that many extant flighted bird species that now appear to be endemic to single islands were previously more widespread⁴. Here we report that the combination of ancient DNA and palaeontological techniques can provide information necessary for conservation management of such a species, the endangered Laysan duck (*Anas laysanensis*).

The Laysan duck is historically known only from the remote, small (370-hectare) island of Laysan in the northwestern Hawaiian chain, where its numbers have varied between 20 and approximately 500 over the past 70 years^{5,6}. Fewer than 150 survived drought conditions in 1993, illustrating the vulnerability of this population.

Both the Laysan duck and the Hawaiian duck, or koloa (*Anas wyvilliana*), were once thought to be derived from stray

migratory mallards (*Anas platyrhynchos*), a perception that has greatly influenced the recovery programme of the Laysan duck^{7,8}. However, the relationships between the three taxa were uncertain^{7,9}. Recent genetic studies have demonstrated that, while the koloa is indeed closely related to mallards, the Laysan duck is very distinct from either (J.R., unpublished data).

Bones of small ducks occur in Late Pleistocene and Holocene deposits in the chief Hawaiian islands^{3,10}. Measurements of the lengths of the main long bones, with sample sizes between 3 and 27, showed the average lengths of the palaeontological specimens to be intermediate between Laysan ducks and koloa (dns). The bones could not be identified from these data because: (1) skeletal morphology is poorly diagnostic in dabbling ducks⁹; (2) other island ducks changed size during the Holocene¹¹; and (3) the bones could represent an extinct taxon distinct from either the koloa or Laysan duck.

To resolve the identity of the subfossil duck bones, we used PCR (polymerase chain reaction) and ancient DNA techniques¹² to extract and sequence DNA

from two variable portions of the mitochondrial control region (133 and 312 base pairs, respectively). Comparison of the sequences (see figure) clearly shows that the subfossil bones match those of the extant Laysan duck, indicating that the species was formerly widespread in the Hawaiian islands.

Late Holocene subfossils of Laysan ducks, including some bones of non-flying juveniles, have been found in habitats varying from near sea level on the islands of Molokai, Oahu and Kauai to formerly forested areas at high elevations (60–1,800 m), far from permanent water, on Maui and Hawaii. Although the species was evidently widely adaptable, it is highly unlikely that Laysan island (maximum altitude 12 m) presents an optimal habitat for it.

The disappearance of the Laysan duck everywhere but on Laysan island is assumed to result from the same human-induced factors (hunting, habitat destruction, introduction of predators and pathogens) that are blamed for the extinction of other taxa native to Pacific islands. Our data justify the reintroduction of the Laysan duck to parts of its former range in the main Hawaiian islands where adverse factors can be controlled. This study demonstrates the value of using ancient DNA and palaeontological information to design recovery plans for endangered species and should provide a model that can be adapted to many other threatened insular species in the Pacific and elsewhere.

Alan Cooper*

Judith Rhymer†

Helen F. James†

Storrs L. Olson†

Carl E. McIntosh*

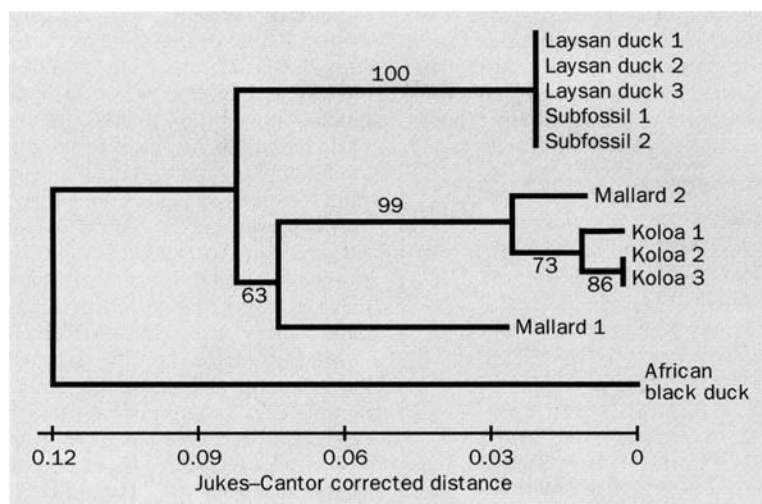
Michael D. Sorenson*

Robert C. Fleischer*

*Molecular Genetics Laboratory,
National Zoological Park,
Smithsonian Institution,
Washington, DC 20008, USA

†Department of Biological Sciences,
Clemson University,
Clemson, South Carolina 29634, USA

‡Department of Vertebrate Zoology,
National Museum of Natural History,
Smithsonian Institution,
Washington, DC 20560, USA



Neighbour-joining two-parameter-corrected distance tree obtained with MEGA¹³. An identical topology is produced by parsimony analysis, with a consistency index of 0.9 (PAUP 3.1.1; ref. 14), and bootstrap values from 1,000 replications are given. The control region sequences were amplified with PCR using primers: C1 (L 00078) 5'-GTATTTGGT-TATGCATATCGTG; C1R2 (H 00390) 5'-CGATTAGTAAATCCATCTGGTAC; C2 (L 00736) 5'-ATCTAAGCCTGGACACACCTG; t-Phe (H 01251) 5'-TGGCAGCTTCAGTGCCATGC; and GCDR (L 01117) 5'-TATTAGAGAAATCCAGTAC, where the strand designation and 3' position in the published chicken sequence are given in parentheses. Sequences were obtained as described¹² from an African black duck outgroup, two subfossils from different caves on Hawaii, and three koloa, three Laysan ducks and two genetically divergent mallards. Considerable genetic diversity has been detected within mallard taxa (J.R., unpublished data), so the two most divergent haplotypes are used in this analysis. The subfossil sequences were extracted and amplified in Washington DC, in a dedicated laboratory in a remote building, before studies were carried out in both DC and South Carolina on extant taxa. There were 366 homologous positions for all taxa, and 36 variable sites within the ingroups. The subfossil sequences clearly group with the Laysan duck, and are genetically distant from the koloa and mallards.

- Milberg, P. & Tyrberg, T. *Ecography* **16**, 229–250 (1993).
- James, H. F. & Olson, S. L. *Ornith. Monogr.* **45**, Pt II. (Am. Ornith. Un., Washington DC, 1991).
- Olson, S. L. & James, H. F. *Ornith. Monogr.* **44**, Pt I (Am. Ornith. Un., Washington DC, 1991).
- Steadman, D. *Science* **267**, 1123–1131 (1995).
- Warner, R. E. *Condor* **65**, 3–23 (1963).
- Marshall, A. P. *Bird Conserv. Int.* **2**, 239–251 (1994).
- Weller, M. W. *The Island Waterfowl* (Iowa State Univ. Press, Ames, 1980).
- Moulton, D. W. & Weller, M. W. *Condor* **86**, 105–117 (1984).
- Livezey, B. C. *Wildfowl* **44**, 75–100 (1993).
- Giffin, J. G. *'Elepaio* **35**, 1–3 (1993).
- Worthy, T. H. *J. Zool.* **215**, 619–628 (1988).
- Cooper, A. in *Ancient DNA* (eds Herrmann, B. & Hummel, S.) 149–165 (Springer, New York, 1993).
- Kumar, S., Tamura, K. & Nei, M. *MEGA: Version 1.01* (Pennsylvania St. Univ., 1993).
- Swofford, D. *PAUP: Version 3.1.1* (Illinois Nat. Hist. Survey, Champaign, 1993).

Quantum incoherence

Daniel C. Dennett

Evolving the Mind: On the Nature of Matter and the Origin of Consciousness. By A. G. Cairns-Smith. Cambridge University Press: 1996. Pp. 329. £16.95, \$24.95.

AFTER decades of persistent work by researchers in many fields, building foundations and patiently filling in details, the gigantic jigsaw puzzle of consciousness is beginning to come into focus. As large assemblies fall into place with a gratifying convergence of details drawn from different disciplines, the pace is quickening. Everybody wants to be in on the delicious task of describing what the Big Picture is going to look like, predicting the outlines before the mopping-up operations confirm them. Well, not quite everybody. There are also those who dislike what they see happening: consciousness is turning out to be 'just' a great big jigsaw puzzle. What? No cosmic revolutions in quantum (or meta-) physics? No Impenetrable Mysteries? Bummer!

What does a theory of consciousness look like to a chemist whose home base is molecular evolution, but who has educated himself strenuously if patchily in cognitive neuroscience and even philosophy of mind? A. G. Cairns-Smith is a brilliant explainer of difficult ideas, bringing to the task an imagination that is magnificently disciplined by detailed scientific understanding. He is also open-minded. His book will tantalize participants and onlookers of all persuasions, for after lucidly describing many of the best pieces of the puzzle, and showing how they could (almost?) complete the picture, he veers off to join forces with the quantum-physics-to-the-rescue squad. This is all the more thought-provoking because he appreciates, and indeed eloquently expresses, the well-known line of reasoning that persuades most of the puzzle-solvers that consciousness will prove to be like all the other dazzling phenomena of life (self-replication, self-transformation, self-repair, self-fuelling, self-protection): explicable in terms of molecular and cellular machinery (and higher-level assemblages of such machinery) without having to invoke any amplification of subatomic weirdness. Has Cairns-Smith seen the importance of something we others have underestimated?

No, I think it is he who has underestimated an opportunity staring him in the face: consciousness is not some further phenomenon occurring in the brain, but is constituted by all the phenomena that individually do not count as instances of consciousness — in the same way that life is not some mysterious phenomenon over



"The Prince of Darkness: Dagol" from a late-eighteenth-century German work on magical arts. It appears in "Abracadabra: The Magic of Medicine", an exhibition at the Wellcome Institute for the History of Medicine, London, from 21 June to 26 October.

and above the components listed above, but is constituted by their ensemble occurrence. What makes his swift dismissal of this prospect all the more tantalizing is that he himself sees the ominous parallel with vitalism — and rejects it: "Indeed, apart from its origin, life, it seems to me, is essentially explained as a phenomenon by a combination of conventional molecular biology and the neo-Darwinian theory of evolution. But consciousness is another matter altogether, one on which molecular biology has so far provided little illumination." One might suppose that Cairns-Smith, like many others, rejects this 'constitutive' option out of hand

because it is so initially counter-intuitive, but he knows better than to do that; he recognizes that a theory of consciousness "should seem crazy (anything evidently sensible would have been established ages ago)". Why, though, does he not heed his own principle and at least give this 'crazy' prospect a serious exploration?

We can mark the moments where the missteps take place. After a brilliant exposition of the tricky relationship between subatomic physics and the biochemists' 'ball-and-stick' models of macromolecules, he proceeds to show how these machines work together to compose greater machines. I have never encountered a clearer or more vivid account of the spectacular ingenuity of cellular design and operation, eventually focusing on the details of "the computers within the Computer" — neurons and their paracrine and endocrine signalling systems.

A crisp, no-nonsense primer then takes us up to the next level, where his analysis of specialized neural circuits (and how they probably evolved) lets his readers arrive, in good company, at the idea that, thanks to the many activities of these specialized subsystems, there is neither a localized destination for the 'inbound' traffic, nor a localized source for the 'outbound' traffic of consciousness. The work to be done by consciousness must be spread out "all over the place" in the brain. This almost perfectly sets the stage for the proposed wedding of neuroscience and phenomenology. What now has to be considered is the initially mind-boggling idea that both sides of consciousness (both the 'given' and the 'taking' of the 'given' with all its repercussions) have to be inextricably intertwined in all these distributed activities. If both the self and the feelings that the self finds nasty or nice are jointly constituted by

the relations and interactions of the neuronal networks, there is no further or left-over appreciation-phenomenon to be explained. So no further feelings have to be generated, somehow, by novel neuronal activities — no task remains for which quantum effects might possibly come in handy.

Cairns-Smith's stage-setting has discouraged this perspective, however, by postulating the need for a "system-3": "Chemical, neuronal, conscious, there are these three forms of control to be discerned." Why is conscious control an additional sort of control, not simply a function of the chemical and neuronal

competences he has already so brilliantly surveyed? Because, he says, "our multi-sensory experience of the world has a certain unity if only because it is so much simpler than the frantic computing that underlies it." He goes on, however, to see that this answer is too strong: "I think that our consciousness is highly but not completely integrated, and that it is more integrated at some times than others", but then he swings back the other way, asserting that "there should be one boss" and eventually concluding: "Our conscious selves seem to be quasi-independent agents which operate through feelings. And I think they are that."

Over and over again, Cairns-Smith asks himself the right questions and even gives the right answers (in my opinion): "Our consciousness has a certain unity and yet at the same time it has passive and active aspects to it: perception and volition. Anyway that is what it seems 'from the inside', and it is what is needed finally to short-circuit a regress at the level of consciousness. Consciousness must do something. Feelings must have effects." Precisely. But if we take that seriously, we should be able to see — however crazy it seems at first glance — why feelings are not anything over and above the patterns of activity of all those neural signalling processes. And again: "We should consider the possibility that feelings have their own inner workings too." Indeed, but why should those inner workings be postulated to involve quantum-level effects? Cairns-Smith follows Henry Stapp (author of *Mind, Matter, and Quantum Mechanics*, Springer, 1993)

in following William James, and quotes as decisive a passage from James, pointing out, correctly, that James's argument is "hardly conclusive.... But I do not think this spoils his general argument." Well, why not? "The kind of unity of brain which is achieved by assembling molecules together to make successively higher-order machines gives us no adequate insight into how conscious experience is so much of a piece." It all comes down to being very impressed with the much-of-a-pieceness of consciousness.

By his own account, then, he strikes only glancing blows — often retracted — against the constitutive option. Having done that, he goes on to give an illuminating survey of the quantum options, pointing out calmly and clearly the extravagances they variously involve. He is so sane, so honest in this undertaking that his book amounts to the best advertisement yet for this family of options. If anybody can wrest a coherent story out of this jumble of spookiness, Cairns-Smith can, but, for just that reason, many of us conservatives will take heart that we are on the right track after all, for although he succeeds handsomely in fending off brusque dismissal, his attempt to show that we need such measures falls well short. He closes with a dialogue between Advo and Krit, an attempt to give the other side its proper innings. How I ached to take over the controls of Krit! □

Daniel C. Dennett is at the Center for Cognitive Studies, Tufts University, Medford, Massachusetts 02155, USA.

Out of the shadows

Sally Temple

Neuroglia. Edited by Helmut Kettenmann and Bruce R. Ransom. *Oxford University Press*: 1995. Pp. 1,079. £150, \$195.

DURING his defence of his thesis, a PhD student recently bemoaned the fact that a major neuroscience textbook devoted roughly two of its thousand pages to glial cells. When I was a PhD student, a favourite professor was touring with a lecture entitled "The Astrocyte: Unsung Star of the Nervous System". In keeping with this view of glial scientists as underdogs, the dedication of *Neuroglia* is: "To those who believed in glial cells during the long, dark period when the neuron concept dominated brain science".

Now the underdog definitely has its day. *Neuroglia* is a comprehensive text that covers the gamut, from morphology to physiology to pathology. It is a thorough, thoughtful assemblage of 69 chapters by prominent researchers in the glial field. In such a large book, differences in style are often apparent, and it is a credit to the editors and contributors that the flow is maintained so well. Overall it is readable, interesting and informative.

There is some patchiness, inevitable in a book of this breadth. Rather than pinpoint the relatively minor gaps, I would rather point out a few highlights such as the chapters on non-mammalian (largely invertebrate) glia, microglia (the under-



THE strong affinities between the developing sciences of linguistics and anthropology early this century are shown by the work of Frances Densmore of the Smithsonian Institution, here seen playing a phonographic record of native American utterances to Mountain Chief, a Blackfoot, who interpreted them. The picture is taken from the new paperback edition of *The Science of Words* by George A. Miller. "Lucid, interesting and original", wrote Paul Fletcher in a review in *Nature* 352, 30 (1991). W. H. Freeman/Scientific American Library, \$19.95, £14.95.

dog of glioscience also has a refreshingly prominent place), gap junctions, extracellular space and gliosis. The authors have, in general, covered their subject from classical times to the present day to represent an in-depth look at glial cells from the accumulated vision of numerous dedicated scientists. I would recommend the book not only to neuroscientists but also to anyone interested in examining the intricacies of specialized cells.

In the study of neurobiology, it seems that glial scientists initially had the harder task. While those studying neurons could point to an obvious function, and focused their attention largely on electrical activity, glial cells did not give up their roles so easily. Researchers had to construct ingenious theories about glial cell function on the basis of a wealth of anatomical and physiological studies. In doing so they may, in the end, have created a more holistic understanding of glial cells compared to neurons. *Neuroglia* is the culmination of nearly a century of work in the shadow of neurons, and, as far as domination by the neuron concept goes, this book, at a mighty seven pounds in weight, goes a long way to redressing the balance. □

Sally Temple is in the Department of Pharmacology and Neuroscience, Albany Medical College, Albany, New York 12208, USA.

Romantic ideals

Graham Farmelo

The Myth of Scientific Literacy. By Morris Shamos. *Rutgers University Press*: 1995. Pp. 238. \$27.95.

TRYING to define scientific literacy is like trying to nail a pudding to a balloon. Should scientifically literate people be expected, unlike most scientists, to be familiar with *every* branch of science? If not, what facts should the scientific literati know? Should they understand the methods of scientific enquiry and how science is administered?

Although the concept of scientific literacy is infuriatingly imprecise, its promotion has become something of a cottage industry in the West. Indeed, few commentators can be heard actually opposing (or even questioning) attempts to improve what is generally known as the 'public understanding of science', another multiply ambiguous and unfortunate catchphrase. The overwhelming consensus among scientists and politicians is that the achievement of scientific literacy is an important strategic goal for countries that wish to compete in the new global economy. Yet it is by no means obvious that the aims of many advocates of science literacy are achievable or that this type of

literacy is essential for the public good.

In this 238-page indictment of sloppy thinking about the public dimensions of science education, Morris Shamos addresses these issues, mainly in the context of his experience in the United States (although most of what he says applies to some extent to other Western countries, notably the United Kingdom). He suggests that it is a fallacy to assume that the achievement of scientific literacy in schools would necessarily lead to a corresponding effect among adults, who have never accepted that they need to know much about science. (You can hardly blame them: most successful people manage perfectly well without it.) His depressing conclusion is that the idea that a notable degree of scientific literacy is achievable among the US public is "little more than a romantic idea".

Shamos is an emeritus professor of physics at New York University and a respected veteran of a wide variety of science education projects in schools and universities. It must, therefore, have been a painfully cathartic experience for him to have written this critique of the arguments that he, or at least many of his colleagues, have used to justify their work over the past few decades. Although his tone is sweetly reasonable throughout, many of his fellow science educators will be shocked by the brutal candour of his analysis and unsettled by his blunt conclusions. For someone who apparently believes that he has spent most of his working life building and maintaining a house of cards, he is remarkably free of rancour and regret.

He assembles his case in a series of chapters that probe and question orthodox thinking about scientific literacy. After questioning whether there really is a crisis in science education, he considers the special nature of science (which he carefully distinguishes from technology and engineering) and reviews the history of initiatives in the public understanding of science. In the crux of the book, he then goes on to discuss the progress of the US 'science literacy movement', from the pioneering work of the philosopher and educator John Dewey in the early part of this century, to the drive for science literacy after the Second World War, the panic that followed the 1957 launch of the Soviet Sputnik and the current disquiet over high-school standards and public attitudes towards science. Despite a slew of well-intentioned educational projects, the proportion of Americans who might reasonably be considered to be scientifically literate is now no more than about six per cent.

Past failures of policy have led to some bizarre initiatives, notably by President George Bush, the 'Education President' as he wished to become known, who set the risibly optimistic target that US students

would lead the world in achievement in science and mathematics by 2000. Comments on its inevitable failure will no doubt be drowned by the millennial ballyhoo — if so, another opportunity to learn from impractical optimism about science education policy will have been wasted.

Shamos is in no doubt that the scientific literacy movement has failed — failed to be clear about what it wants to achieve, failed to set realistic targets and failed to make a significant impact even in its own wishy-washy terms. The clear implication is that vast sums of money have been frittered away on science literacy projects and that although rationality, intellectual rigour and self-criticism characterize good scientists, these qualities do not yet apply to most science educators. As one embarks on the section entitled "Do educational experiments ever fail?", one braces oneself for another Shamos onslaught. It duly arrives, with the expected pain.

Having wreaked havoc with the conventional agenda of scientific literacy, he does the decent thing and suggests another perspective, although without much enthusiasm. He suggests that we should define scientific literacy in terms of goals that just might be achievable: it might, for example, be possible to improve the public's awareness both of science and of what can be expected from it. Also, much more could be done to give laypeople a clearer, better-informed voice in the scientific enterprise. Shamos uses these and other ideas to suggest the basis of a new approach to the teaching of science to general (that is, non-science) students, suggestions that will affront the beleaguered hard-liners.

It is possible to carp at Shamos's pedestrian prose, his prolixity and his repetitiveness, but there is no denying that he has marshalled a formidable case against decades of muddled thinking about scientific literacy. No one concerned about the science education agenda can be indifferent to his arguments: some will reflect that they have a good deal of rethinking to do, others that they have a lot to answer for. □

Graham Farmelo is at the Science Museum, Exhibition Road, London SW7 2DD, UK.

New in paperback

Einstein, History, and Other Passions: The Rebellion Against Science at the End of the Twentieth Century by Gerald Holton Addison-Wesley, \$14. A distinguished scientist and historian of science draws on the life and work of Einstein to examine the role of science in Western culture and argue for the necessity of reason.

Legal protection

Susan E. Foden

Patent Strategy for Researchers and Research Managers. By H. Jackson Knight. Wiley: 1996. Pp. 166. £32.50.

THE aim of this book is to encourage the "technical professional" to think strategically about filing patent applications and building patent portfolios. The author not only describes the basic concepts of intellectual property and patent procedure, but also considers the human elements of the process — the driving force behind invention and the personal interactions between researchers, managers and patent agents at various stages of the research process.

The first chapter provides a sound introduction to patent law, the workings of patent offices and requirements for obtaining a patent. Particularly informative are the sections on types of patents and on the parts of a patent. In the latter, the author clearly explains the various reference criteria for patent documentation — important information often overlooked in other texts on preparing a patent proposal.

The value of a patent lies mainly in the scope of exclusivity that the owner may hope to secure. In outlining several easy-to-follow models of how this may be best achieved, the author begins to put the whole patenting business into a realistic perspective in terms of its value and exploitation. The passing references to licensing seem out of place, however: licensing and licensing strategy are enormous topics in their own right, and it is impossible to do justice to them without looking more widely at the area of exploitation opportunities. Indeed, both here and elsewhere in the book, it is not clear what area of science the author is talking about, or to whom he is talking. The relevance of his points often depends on whether the reader is a private inventor, a university researcher, a start-up company manager or an employee of a large international corporation. The author's background is the last of these, and on the whole he seems to be writing with this environment in mind. Those whose experience is mainly with inventions arising in the academic sector and in areas of rapidly moving and highly competitive biomedical research will find the book less useful.

But the real meat of the book is the thorough, well-presented third chapter on developing a patent strategy. The author guides us through fundamental thinking that could well apply in a variety of circumstances. He deals with the military principles behind strategic thinking and goes on to the development of a patent strategy model. Several summary tables

offer useful checklists or reminders for later reference.

After exposure to the strategic thinking mode, the following chapter on global filings reads like a harsh return to the facts and practicalities introduced earlier in the book. My own preference would have been to proceed directly to the next chapter, "Researching with Intellectual Property in Mind", on the human elements of the process. The author starts with debatable statement: "Researchers, as a whole, are not thought of as inventors". Nevertheless, it is interesting to be taken through the more reflective aspects of inventive philosophy and to look at the environments and management styles that may encourage — or discourage — the process. There naturally follows a chapter that looks in some detail at the relationships between inventors, patent attorneys, patent agents and patent liaison officers.

Thereafter the book tends to lose its way, dealing as it does with disclosure and filing decisions and the responsibilities of the researcher after the filing. Repetition here of earlier material on the patent process and procedures makes for somewhat dreary reading. And far from being given succinct or comforting conclusions, the reader is left dangling among non-strategic issues such as maintenance fees and prosecution decisions.

Overall, the book offers a readable description of the patent process and a solid approach to the issues of filing and strategy. But do bear in mind that it is biased towards nonbiological research in a company environment. □

Susan E. Foden is at Cancer Research Campaign Technology Ltd, Cambridge House, 6–10 Cambridge Terrace, Regent's Park, London NW1 4JL, UK.

New Journals

This year, *Nature's* annual New Journals review supplement will appear in the issue of 5 September. Publishers and learned societies are invited to submit journals for review as well as details of any eligible electronic journals. Journals that first appeared during or after June 1994 and issued at least four separate numbers by the end of May 1996 will be considered. Frequency of publication must be at least three times a year and the main language used must be English. Deadline for submission is 14 June.

When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates. For further information please contact Peter Tallack, *Nature*, Macmillan Magazines Ltd, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567; e-mail p.tallack@nature.com.

Delicate investigations

Stephen Mason

Lines of Light: The Sources of Dispersive Spectroscopy, 1800–1930. By John C. D. Brand. Gordon and Breach: 1995. Pp. 266. £84, \$140 (hbk); £36, \$60 (pbk).

FOR Faraday, light constituted "a most delicate investigator of molecular condition". In *Lines of Light*, the distinguished spectroscopist John Brand relates how Faraday's perception was realized historically in the development of optical spectroscopy. The book covers instrumental innovations, theoretical interpretations and chemical and astrophysical applications, from the time of Fraunhofer (1787–1826) to the symbiotic growth of quantum mechanics and spectroscopy during the decades between the two world wars, before the advent of the laser.

Each of the 12 chapters deals with a particular topic and period and ends with brief summaries of the lives of the principal workers. After providing a general introduction to physical theory during the nineteenth century, the author begins with the transverse-wave theory of light and the extension of the known spectrum to the infrared and ultraviolet ranges. He gives an account of the evolution of methods and standards of wavelength measurement and describes emission and absorption studies, including laboratory and stellar spectrum analysis. Following a description of pioneering infrared spectroscopy and matter–radiation equilibrium problems, the author discusses the mapping and analysis of atomic spectral series and the molecular spectra of gases, leading on to the empirical generalizations and early quantum interpretations. He ends with the salient discoveries of atomic and molecular electron orbitals, the Raman effect, the Franck–Condon principle, and electron and nuclear spin — all part of the convergence of theory and experiment that resulted in the alliance of spectroscopy and quantum mechanics.

Some 36 illustrations add to the clarity of the text. As the first substantial history of spectroscopy, this book will appeal to both students and historians of physical science. □

Stephen Mason, emeritus professor of chemistry at the University of London, is at 12 Hills Avenue, Cambridge CB1 4XA, UK.

Correction. In Patrick Naylor's review of *Pattern Recognition and Neural Networks* by B. D. Ripley (*Nature* **381**, 206; 1996), the publisher was incorrectly given as MIT Press, instead of Cambridge University Press. Our apologies.

The history of the galaxies

M. Fukugita, C. J. Hogan & P. J. E. Peebles

Astronomical observations now reach far enough back in time, in enough depth and detail, to reveal the history of galaxies since their formation. The early Universe contained a network of gas clouds that filled much of the space between the young galaxies, where stars were forming at a high rate. Since then, intergalactic space has been swept clean, and galaxies have continued to convert the dwindling supply of gas slow into stars.

ACCORDING to the relativistic Big Bang model¹, the Universe started with a uniform mass distribution, but was gravitationally unstable to the growth of clustering. The present concentrations of matter in galaxies and systems of galaxies grow out of small departures from homogeneity present in the very early Universe. Galaxy formation is much too complicated for a detailed analysis from first principles, so we attempt to piece together a picture of what happened from the observational evidence. For many years, the main elements of the picture depended on the fossil evidence in relatively nearby galaxies, including our own. But recent advances in astronomical imaging and spectroscopy have added a new dimension to the search for an understanding of how galaxies formed: we can see what happened over a considerable span of cosmic time. Observations now penetrate to such great distances that the light travel time from the source of the radiation to the present is comparable to the expansion time in the Big Bang cosmology, allowing a direct view of cosmic evolution from protogalactic gas clouds, to young galaxies, to the mature systems of the present epoch.

The light from distant objects carries information in spectra as well as images and from absorption processes as well as emission. The most direct visualization comes from images, the deepest now being the new Hubble Deep Field (see cover; also available at <http://www.stsci.edu/ftp/observer/hdf/hdf.html>). These images show starlight in emission, allowing measurements of brightness, broadband colour and morphology, and conveying qualitative information about the amount of star formation and spatial organization in early generations of galaxies. Dispersion of light into 1,000 or more colours in the images of brighter galaxies allows a more detailed study of the emitting stars and gas, and especially, through the measurement of redshift, their distances. A few extremely distant quasars—the rare, bright, active nuclear regions of some early galaxies—are so luminous that their light can be dispersed into over 30,000 colours; at this resolution, line absorption by foreground gas reveals a detailed one-dimensional map of galactic and protogalactic gas all the way back along the line of sight to the quasar, through the last three-quarters or more of the history of the Universe.

The directly measurable coordinate along the line of sight is not time but redshift (z); the expansion of the Universe stretches the wavelength of the light from a distant galaxy by the factor $1+z$ if the Universe has expanded by the factor $1+z$ since the light we detect left the galaxy. Thus, measurements of the redshift date the slice of cosmic history that we explore by looking out in space and back in time along our past light cone—like the history recorded in an ice core, a tree-ring sample, or geological strata surrounding a dinosaur footprint, except that the astronomers' record is imprinted on the images and spectra of light arriving from distant objects. The character of the emission spectrum contains information on the source of the light, whether from a young or an old star population; the star formation rates in individual galaxies can, for example, be estimated from the fluorescent line emission from gas excited by light from hot young stars² or the shape of the ultra-

violet continuum³. In the case of quasar absorption spectra, the light from the quasar has passed through many different gas clouds at different redshifts, and each cloud has left in the observed spectrum an imprint of absorption lines at wavelengths determined by the redshift of the cloud. The absorption lines from various ions carry information about the internal velocity, temperature, column density, composition and ionization state of the intervening gas.

Our purpose here is to review the progress in these types of observation and propose a synthesis of the results in a picture that contains many of the trends one would look for in a Universe that is evolving from a dense, nearly homogeneous early state: the clearing of intergalactic material as the gas is concentrated into galaxies, the emission from large numbers of newly formed stars in young galaxies, the change in composition of material from a nearly primordial mix dominated by hydrogen and helium to one enriched in heavier elements, the growth of galaxy clustering with time, and the changes in galaxy morphology as these objects approach their present slowly evolving state. We trace the evolution back in time from the present to what seems to be the epoch of formation of the giant galaxies, in the redshift interval $z = 1-3$ where both imaging and spectroscopy now yield abundant detailed information.

Galaxies at $z \leq 1$

Giant galaxies today—those with luminosity similar to our own Galaxy—usually evolve only slowly, on time-scales comparable to the expansion time of the Universe (~ 10 Gyr), although as we will discuss there are examples of rapid evolution. Present-day galaxies serve as the baseline for comparison to the younger-looking, more rapidly evolving objects observed at redshifts greater than unity.

Giant galaxies are often divided into two major categories, spiral and elliptical. The visible light of spiral galaxies comes mostly from disks of stars on nearly circular orbits. Stars are forming from gas in the disks, but the rate is usually slow because the mass fraction in gas is low and star formation tends to be self-limiting: formation of stars heats the gas and prevents further small-scale gravitational collapse. Elliptical galaxies contain almost no neutral or molecular gas. The ageing stars in an elliptical galaxy orbit more isotropically than the younger stars in a spiral, filling a triaxial (roughly ellipsoidal) shape. Ellipticals are denser than spirals, and tend to be found in denser environments. These simple categories are not sharply defined, however: spiral galaxies often have an elliptical-like spheroid component of old stars, ellipticals sometimes contain clouds of gas and dust and young star clusters, and the intermediate S0 galaxies look like spirals that have been stripped of gas.

In the high-angular-resolution images from the Hubble Space Telescope (HST) the giant galaxies at $z \approx 1$, both ellipticals and spirals, are somewhat bluer than their present-day counterparts, which is evidence of greater star formation rates, and spiral disks look a little fuzzier, yet to be settled into the prototypical

FIG. 1 The M81 group, at 3.5 Mpc distance, shown here in a conventional optical image (right) and in H I emission (left)¹⁴. The outer H I contour is at $1.8 \times 10^{20} \text{ cm}^{-2}$, corresponding to mildly damped Lyman- α absorption. The images are 100 kpc square at the distance of M81.

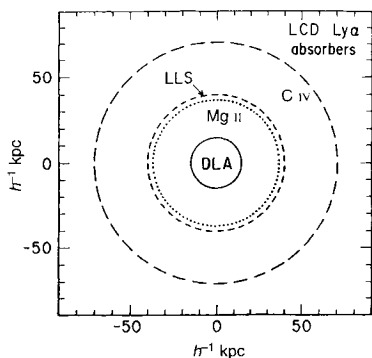
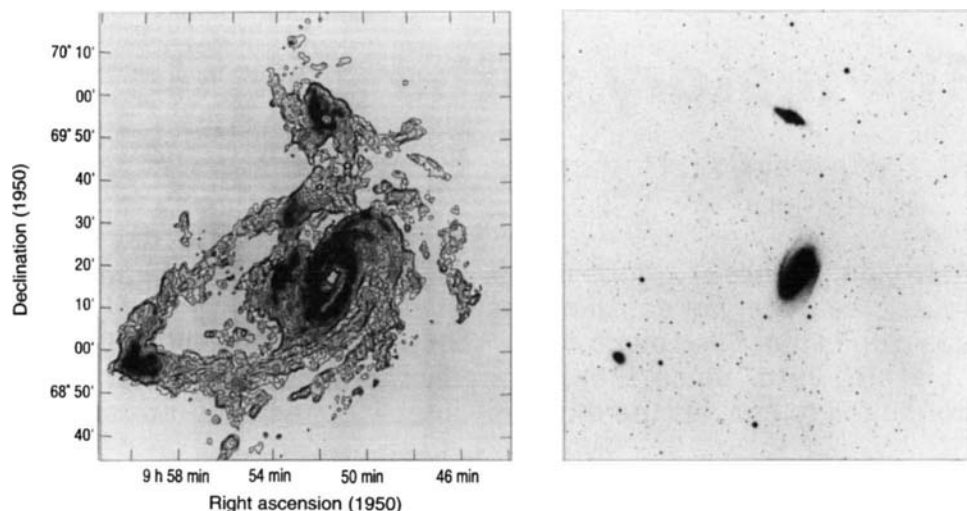


FIG. 2 The absorption profile of a typical galaxy at $z \approx 1$, from refs 12 and 22. The circles show the area within which a line of sight causes absorption by the species labelled: DLA, H I damped Lyman- α ; Mg II, Mg II absorbers; LLS, H I Lyman-limit systems; C IV, C IV absorbers; and LCD, low-column-density H I. This sketch does not attempt to indicate the texture of the halo, but in spite of its simplicity it properly indicates that the cross-section has to be close to filled and circular to account for the small scatter in the maximum impact parameter for detection of absorption features. Even the scales of the various zones vary surprisingly little from galaxy to galaxy, and also significantly, nearly all absorbers of these types are found within this vicinity of a giant galaxy. The LCD zone extends to about 200 kpc for H I column density thresholds of the order of 10^{15} cm^{-2} .

morphology, but they otherwise look much the same as they do today⁴⁻⁶. Consistent evidence comes from a deep redshift survey of galaxies selected in the near-infrared (the I band centred on wavelength $0.85 \mu\text{m}$). The detected light is emitted in the blue (taking account of the redshift), and probably comes from a mix of stars similar to what is seen in present-day normal galaxies. The counts of galaxies as a function of brightness and redshift are about consistent with what would be expected from the properties of the nearby sample of galaxies^{7,8}. The lack of rapid evolution to $z \approx 1$ or even higher z also is indicated in the deep galaxy counts in the infrared K band ($2.2 \mu\text{m}$)⁹⁻¹¹ and in the giant galaxies identified from quasar absorption line studies¹². That is, the evidence is that the giant galaxies already were mature slowly evolving systems at $z = 1$.

In certain circumstances, however, these giant galaxies evolve faster. The equilibrium in the disk is disrupted on occasion by interactions between galaxies: the tidal interaction triggers flows of gas and angular momentum, leading to rapid conversion of gas to stars at rates well above the sustainable equilibrium, often resulting in starburst galaxies^{3,13}. Such galaxies are usually very

bright in infrared emission from dust produced along with the burst of young stars: the dust absorbs and reradiates most of the light from the massive young stars. The energy deposited by the stars in the interstellar gas may eventually quench the starburst and leave some diffuse gas in the galaxy, or it may drive most of the remaining gas out of the system.

Even in the Universe today there are examples of the central process of galaxy formation, the exchange of gas between galaxies and the intergalactic medium (Fig. 1). The sweeping streams¹⁴ in the distribution of the gas around the central three galaxies in the nearby M81 group suggest the gas was swept out of the galaxies, perhaps mixing with a component falling in from greater distances. The stars in the galaxies are likely sources for the heavy elements in these gaseous haloes, and the halo gas may be falling back into the galaxies and contributing to star formation. But neither starbursts, nor the exchange of gas with the intergalactic medium, can have had much effect on the evolution of most giant galaxies at redshifts $z < 1$, because these objects show so little difference from today's galaxies.

The slow evolution of the giants, however, is not the whole story: gas-rich galaxies, particularly the less luminous ones, were much more numerous than today even at modest redshifts. There were two early indications of substantial evolution at modest redshift: rich clusters at $z \approx 0.5$ contain a larger proportion of gaseous star-forming systems than is observed nearby¹⁵, and counts of galaxies selected in blue light (that is, light emitted in the ultraviolet) are larger than would be expected if the galaxy population were not evolving¹⁶⁻¹⁸.

There is no generally accepted interpretation of all the details of the deep galaxy counts, but it is known that there are significant contributions to the excess blue counts from galaxies that are somewhat less luminous and bluer than is typical of the giants (ref. 8, and R. S. Ellis, T. J. Broadhurst, M. M. Colless, J. S. Heyl and K. Glazebrook, manuscript in preparation), and more irregular in form, as seen in HST images⁴⁻⁶. The colours, irregular forms and decreasing abundance with decreasing redshift suggest that we are seeing galaxies made bright in the ultraviolet by massive short-lived stars that are rapidly exhausting the supply of gas for new generations of stars^{4-6,19}. This suggests that star formation in these galaxies has started only at $z \leq 1$.

This picture of delayed star formation in less-luminous galaxies is tested by studies of the spread of ages of stars in present-day galaxies. There is evidence that the stars in the halo of the Milky Way galaxy have a small scatter in ages²⁰, as would be expected if the spheroid in our Galaxy formed along with the elliptical galaxies at high redshift. The distribution of star ages in nearby extreme dwarf galaxies is seen to be broader, consistent with late and non-coeval formation and evolution in lower-mass systems.

Diffuse gas at $z < 1$

We have a remarkably detailed picture of the distribution of gas clouds outside the galaxies from the spectra of quasars: clouds that happen to lie along the lines of sight produce absorption lines at wavelengths determined by the cloud redshift or distance. This gives us an unbiased census of diffuse atomic hydrogen and heavier elements wherever the gas may be.

At redshifts $z \lesssim 1$ gas clouds with relatively high column density, $\Sigma_{\text{HI}} = (3 \times 10^{17})$ to (3×10^{21}) neutral atomic hydrogen atoms per cm^2 , invariably are in the vicinity of a galaxy, at a radius that depends on the cloud column density, the atom and its ionization^{21,22}. At $z \approx 1$, most giant galaxies have the halo of absorbing clouds shown schematically in Fig. 2. For example, clouds with $\Sigma_{\text{HI}} \approx 10^{17} \text{ cm}^{-2}$ (the so-called Lyman limit at which the gas is just optically thick to ionizing radiation at the threshold) are found only within about 40 kpc of a giant galaxy²². Apart from a weak dependence on mass, this absorption profile is independent of almost all properties of the galaxy, suggesting a connection with infall rather than stellar-driven outflow, and is remarkably close to spherical in character, suggesting a quasi-static, pressure-regulated process rather than a free ballistic infall. The gas clouds shown in Fig. 1 have column density $\Sigma_{\text{HI}} \gtrsim 10^{20} \text{ cm}^{-2}$, which is well above the Lyman limit, and are larger in extent than is typical of present-day galaxies, but they illustrate the point that gas clouds near the Lyman limit tend to be very near galaxies.

At lower column densities, $\Sigma_{\text{HI}} \approx 10^{14} \text{ cm}^{-2}$, the clouds still tend to be within 200 kpc or so of a galaxy²³. That is, most of these clouds populate regions which occupy only one part in 10^3 of the volume of space. At the present epoch, and at the lowest observable column densities, $\Sigma_{\text{HI}} \approx 10^{12} - 10^{13} \text{ cm}^{-2}$, the gas clouds avoid the immediate neighbourhoods of galaxies, but they also tend to avoid the broad empty regions between the concentrations of galaxies (refs 24–26, and J. M. Shull and S. Penton, personal communication).

In summary, at the present epoch and back to redshift $z \approx 1$, baryons in clouds left over from galaxy assembly have been quite efficiently swept from the voids between the concentrations of galaxies, leaving the galaxies with gaseous haloes that may be contributing to the present generally leisurely rate of star formation.

Mean densities of baryons in stars and gas

Estimates of the amounts of baryons (the neutrons and protons of familiar matter) present in various forms in the Universe today are useful for comparison to the amount observed in the more rapidly evolving systems at higher redshift. There is evidence that we have a nearly complete survey of where the baryons are: the amounts seen in stars, cool gas clouds, and hot clouds of plasma add up to about what is needed for the theory of light-element production in the early hot Universe.

It is convenient to express the baryonic mass in form i as the ratio Ω_i of the mean mass density in this form to the critical total mass density in a universe that is expanding just at escape velocity. (We use the distance scale parameter $h \approx 0.7$, where the Hubble constant is $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$. The uncertainty in this parameter is not the main contributor to the uncertainties in our estimates of the Ω_i .)

Table 1 lists estimates of the density parameters Ω_i . The star masses in spheroids (elliptical galaxies and bulges of spiral galaxies), disks and irregular galaxies are estimated from models of the stellar populations. The relatively little neutral gas in the Universe today is found almost entirely in the disks of spiral galaxies; atomic neutral gas contributes $\Omega_{\text{gas}} h \approx (2.3 \pm 0.6) \times 10^{-4}$ (refs 27, 28), and molecular hydrogen a similar amount. Hot plasma dominates the observed baryon content of rich clusters of galaxies (gas/total $\approx 0.05h^{-3/2}$)^{29,30} and is detected also in some groups^{31–33}.

Most giant galaxies occur in groups which can harbour a significant number of “dark” baryons, either in the form of plasma or compact baryonic objects. Plasma temperatures in

TABLE 1 Baryon densities at the present epoch

Form	Ω_i
Stars in spheroids	0.0025
Stars in disks	0.0015
Stars in irregulars	0.0002
Neutral atomic gas (H I, He I)	0.0005
Plasma in clusters and groups	0.001
Dark baryons in groups	0.003*
Cool intergalactic gas clouds	0.001–0.003
Sum	0.01
Ω_b from light-element production	0.01 or 0.03

Ω_i , density parameter; Ω_b , total baryon density; Ω_{spheroid} , density in stars in elliptical galaxies and spiral bulges.

*Based on assuming universal $\Omega_{\text{spheroid}}/\Omega_b$; actual number may be higher.

groups are lower than in clusters, consistent with the fact that the galaxies move more slowly, and consequently the plasma in many groups may be too cool to be detected directly as a source of X-rays. Compact baryonic objects (MACHOs) include planet-sized systems like Jupiter, stars with masses near or below the limit for sustained nuclear burning, and star remnants. Their abundance is now being probed in our Galaxy by microlensing searches³⁴. We have entered a number in Table 1 which assumes that the number of baryons per spheroid star mass is universal, the same in groups as in the rich clusters, as would be expected if most spheroids formed early, when conditions everywhere were still much the same. The number in Table 1 for baryons in cool low-surface-density clouds is from Shull *et al.*²⁶, with the smaller number more plausible if the clouds tend to be wispy rather than equally wide in all dimensions.

We have a measure of the total baryon density, Ω_b , from the relative amounts of the light elements produced in the hot young Universe³⁵. Acceptable fits to the light-element abundances follow if the baryon density is in the range $0.01 \lesssim \Omega_b h^2 \lesssim 0.02$. There is, however, an argument based on the high deuterium abundance found in some quasar absorbers^{36–38} that Ω_b is as low as $0.005h^2$, or $\Omega_b \approx 0.01$ if $h \approx 0.7$.

The lower number, $\Omega_b \approx 0.01$, agrees with the total in the components listed in Table 1, which allows room for dark baryons only in an amount comparable to that in each of the other major components, perhaps $\Omega \approx 0.003$. If the total baryon density is at the high end of the range of estimates from light-element production, $\Omega_b = 0.04$, the dark baryons must dominate over visible stars by an order of magnitude, which certainly is possible but requires special arrangement. In particular, if mass were sequestered in dark star remnants one would look for considerable debris from the star deaths. The debris could be in smoothly distributed plasma in the voids, but we will argue later that gravity probably swept the voids clean of plasma as well as galaxies and gas clouds. We do not need to invoke any mechanisms to conceal baryons if total baryon density is $\Omega_b \approx 0.01$. As we will discuss, this is comparable to the amount of baryons identified in gas clouds at larger redshifts.

Galaxies at redshifts $z = 1-3$

Although observations of galaxies and possible progenitors of galaxies at $z \gtrsim 1$ are usually biased towards objects whose activity makes them particularly noticeable, we do know that there are objects at high redshift that are quite unlike galaxies at the present epoch, and they are detected in numbers large enough that we must be seeing a major phase of star formation in galaxies. Thus a substantial part of the whole history of the galaxies is observationally accessible, at $z \lesssim 3$. We now discuss some examples of what appear to be young galaxies and the gas clouds that are progenitors of the disks of spiral galaxies.

Optical searches for the identification of radio sources have led

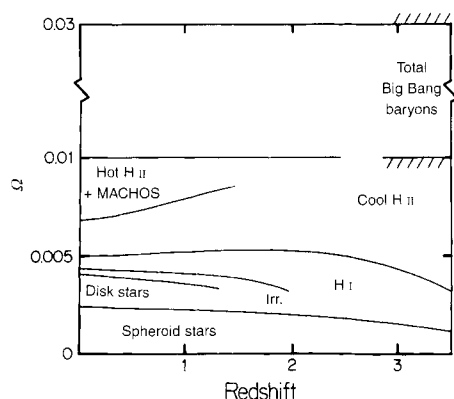


FIG. 3 History of the baryons. The vertical axis is the mass fraction Ω_b , relative to the mass density of the Einstein-de Sitter universe, in various forms of baryonic matter. The history commences at redshift $z = 3.5$; time increases towards the left to the present epoch at $z = 0$. The total baryon mass density probably lies between 0.01 and 0.03, the range inferred from light-element production in the early Universe; the Hubble constant is $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

to the discovery of young galaxies at redshifts above unity³⁹. The most distant galaxy discovered as a radio source is at redshift $z = 4.25$ (refs 40, 41). The continuum radiation emitted at wavelength $4,200 \text{ \AA}$ (and detected at $2.2 \mu\text{m}$) in this galaxy is thought to be starlight. If so, it is estimated from the colour that the stars already were about 0.5-Gyr-old at the time of observation and that the star population would fade to about the luminosity of a large elliptical galaxy at the present epoch⁴¹. The image in what was at emission ultraviolet light is aligned with the radio lobe axis in a way not commonly seen in low-redshift radio galaxies. This pattern repeats in other high-redshift galaxies discovered as radio sources: the images at optical wavelengths in emission tend to look like giant elliptical galaxies, while the alignment in images in ultraviolet light and in gaseous emission suggests the radio activity is still rearranging gas in and around the galaxy⁴². At least one weak radio galaxy has been found at $z = 1.55$ with a spectroscopically measured stellar population age of 3.5 Gyr, implying a much earlier formation redshift⁴³.

Surveys of spectra of faint field galaxies reveal objects at $z > 1$ with high star formation rates, as deduced from the intensities of prominent gaseous emission lines, and irregular images in the optical in emission, which is suggestive of clumpy and very rapid star formation⁴⁴. A substantial subset of such star-forming galaxies, selected by examining colours for the tell-tale zero flux of starlight beyond the Lyman limit, extend beyond $z = 3$ (ref. 45, and M. Giavalisco, C. C. Steidel and F. Machetto, manuscript in preparation). In most cases the luminosity profiles of these very-high-redshift galaxies resemble modern elliptical galaxies (with half-light radii of only 1 or 2 kpc), but the flat spectral energy distribution is characteristic of a much younger stellar population. Indeed the spectroscopic properties (including interstellar absorption features of heavy elements) resemble in remarkable detail those of the prototypical nearby starburst galaxy NGC4214, and the measured velocity dispersions ($180\text{--}320 \text{ km s}^{-1}$) reproduce the range of modern giant ellipticals. The star formation rates in such galaxies, about ten solar masses per year, is enough to produce the number of stars in a giant galaxy if it continues for a Hubble time—a few billion years at that redshift. The number density of such star-forming galaxies is enough to account for substantial fraction of present-day giant elliptical galaxies.

Thus the evidence is that the giant ellipticals formed as galaxies of stars at redshift $z = 3$ or earlier. The close family resemblance of the spheroid components of spiral galaxies to ellipticals suggests the spheroid components also formed at $z \gtrsim 3$ too. As we now argue, it appears that the thin disks of the spirals were added later.

A hint about the formation of the disks of spiral galaxies comes from the quasar spectrum absorption lines produced by gas clouds along the line of sight. The clouds with the largest column densities in neutral hydrogen, $\Sigma_{\text{H I}} \approx 10^{21 \pm 1} \text{ cm}^{-2}$, are prominent because they display absorption in the broad radiation-damped wings of the Lyman- α transition, and hence are called damped Lyman- α clouds. Wolfe⁴⁶ has long emphasized that these clouds look very much like young galaxies, or gas-rich disks in the process of contracting to the present-day thin disks of stars in spiral galaxies.

The pattern of absorption-line redshifts of the low-ionization material in damped Lyman- α clouds is consistent with rotation in disks at circular velocities quite similar to the velocities in present-day spiral galaxies, even to $z = 4.4$ (refs 47, 48). The higher-ionization C IV lines do not show the systematic rotational asymmetry, but the higher ionization suggests the C IV is on the outskirts of the protogalaxy, where the matter is more exposed to intergalactic ionizing radiation. Thus it is reasonable to imagine that the C IV is in clouds moving on disorganized orbits. The sizes and velocity dispersions inferred for the C IV haloes are consistent with those of spiral galaxy haloes⁴⁹.

The total hydrogen mass in damped Lyman- α absorbers at $z \approx 3$ is $\Omega_{\text{H I}} \approx (0.0015\text{--}0.003)h^{-1}$ (refs 46, 47, 50, 51), (or larger if dust in some clouds obscures the image of the quasar⁵²), comparable to what is seen now in stars (Table 1). The mass in the clouds decreases at lower redshift, as if the H I were being consumed between $z = 3$ and $z = 1$.

The presence of heavy elements shows that stars have been forming and dying in these clouds even by $z = 4.4$ (ref. 53). The heavy-element abundances at $z = 2\text{--}3$ tend to be about 10% of the solar abundance typical in the disk of our Galaxy, but with a spread of at least an order of magnitude either way: some have been enriched to about solar while others are still relatively free of heavy elements at about 0.01 times solar^{54–56}. This large scatter is indicative of non-coeval disk formation taking place just around this redshift, because the heavy-element abundance in a closed system reaches the asymptotic value about 1 Gyr after star formation begins. The most enriched material is suitable for modern young stars: the gas would produce stars with the compositions in the thin disk stars of spirals. The less-enriched clouds may be contributing stars to the thick disk component and receiving heavy elements shed by stars in the thick disk and spheroid components⁵⁷. The pattern of element abundances in the clouds often (if not always) hints at that expected for enrichment by the first generation of stars, with relatively larger amounts of oxygen and silicon (produced by rapidly evolving stars) than carbon and iron (produced by slower ones), compared to the interstellar medium today. The same pattern is seen in old stars in our Galaxy^{58,59}.

Because the largest H I column densities in absorption are about one-tenth of the typical accumulated total mass column densities in stars in the disks of present-day spirals, these damped Lyman- α clouds could not be converted to stars to form spiral disks without further rearrangement of the material. Some thin disk stars may already be in place in these clouds, and as gas is converted to stars more may be settling into the clouds. One such cloud at $z \approx 3$ has been observed in a deep image that reveals the presence of stars⁶⁰; they may be early generations in the disk, or a spheroid component, or perhaps a close companion.

If, as is generally thought, the higher-column-density clouds at $z \approx 3$ are associated with galaxies, then the cloud positions ought to be clustered in space, as are galaxies. The observed clustering is even stronger than in present-day galaxies^{61–64}. We return to this result in the last section.

To summarize, the evidence is that the disks of the giant spiral galaxies were assembled at redshifts in the range $z = 3\text{--}1$. We have a reasonable candidate for the disk progenitors, in the high-column-density damped Lyman- α clouds. There is evidence for rapid star formation at $z = 2\text{--}1$, much of which we can ascribe to disk formation, and a corresponding decrease in the neutral hydrogen mass in the damped Lyman- α clouds⁶⁵. The close

resemblance of the spheroid components in spirals to elliptical galaxies leads us to suggest the spheroids formed with the ellipticals, at $z \gtrsim 3$, the present-day disks being added later. The evidence for high circular velocities in the high-column-density clouds suggests the massive dark haloes of the spirals were attached to the spheroids before or along with the damped Lyman- α clouds. It is a suggestive coincidence that the space density of bright optical quasars also reached a peak^{66–68} sometime around $z = 2–3$. Perhaps the high quasar activity is a by-product of the first assembly of giant galaxies, the addition of their disk components, and the accompanying massive starbursts.

Intergalactic gas at redshifts $z = 1–3$

As we have discussed, intergalactic space between the concentrations of galaxies is quite thoroughly cleared of gas clouds at the present epoch. The situation is very different at high redshift: a foam-like network of low-surface-density clouds fills space with mean baryonic mass density comparable to what then is in the damped Lyman- α absorbers.

The low-surface-density clouds, at neutral hydrogen column densities in the range $10^{13} \lesssim \Sigma_{\text{H I}} \lesssim 10^{17} \text{ atoms cm}^{-2}$, appear in quasar spectra as a dense set of Lyman- α resonance absorption lines at redshifts less than that of the quasar—the so-called Lyman- α forest⁶⁹. The neutral hydrogen column densities in the forest clouds are too small to shield against the ionizing radiation from quasars and young stars, so the material is mostly ionized, with neutral fraction of about 10^{-4} at the lowest columns, increasing in the higher-column clouds⁷⁰. At $z = 3$ the mass in plasma in the forest is comparable to the neutral mass in the damped Lyman- α absorbers identified as protodisks of spirals⁷¹. Thus the total mass in the low and high column density absorbers at $z = 3$ is comparable to the total seen in all components at the present epoch (Table 1); the reasonable interpretation is that the two classes of absorbers at $z = 3$ are major contributors to the present-day components listed in the table. Figure 3 sketches our impression of the evolution of the mass fractions Ω , in the major forms of baryons as functions of redshift.

Absorption by the forest gas in the equivalent resonance line of singly ionized helium also appears to be ubiquitous^{72,73}. The fact that a considerable fraction of the helium has been only singly ionized tells us the ionizing radiation does not have a very hard spectrum, especially at high redshifts—possibly due to the effect of gas absorption.

We have a coherence length for the mass distribution in the forest clouds from the comparison of spectra of quasars that are close together on the sky (either gravitationally lensed images of a single quasar or two different quasars). The forest absorption patterns are almost identical when the sightlines are closer than about 10 kpc, still similar at 100-kpc separation, and dissimilar at separations much larger than 100 kpc (refs 74–77). This could mean either that the cloud sizes are about 100 kpc or that smaller clouds cluster on this scale (with a large two-dimensional covering factor); numerical models (see below) suggest that the coherence length is associated with the long dimensions of wispy clouds. The mean distance between clouds along a line of sight is 500 kpc to 1 Mpc, depending on the cosmological model. As this is not much larger than the coherence length, the cloud network roughly fills space. Thus it is not surprising that forest clouds do not show strong clustering along the line of sight⁷⁰: there is not much room for the clouds to cluster.

There are heavy elements even in the forest material^{70,78}. The line strengths are consistent with an abundance about 1% of solar, and display little scatter among most clouds where it has been measured. This points to an early history of low-level star formation, presumably in star clusters small and numerous enough for stellar ejecta to have become uniformly mixed within the clouds. The efficiency of the early star formation must have been low to leave so much diffuse gas at $z = 3$; possibly the gravitational potentials of the first systems were too shallow to withstand the first stellar heating, so the mixing was accompanied by disruption.

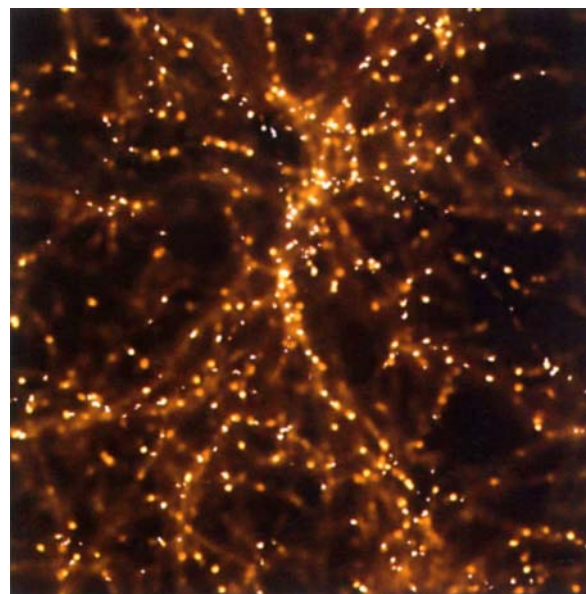


FIG. 4 Visualization of gas responsible for the Lyman- α forest: an image of the neutral hydrogen density in a simulation, produced in a 20-Mpc cube at redshift 2.5, from ref. 84. Low-column clouds resemble a space-filling froth; the dense knots where bubbles meet are high-column absorbers, where gas is turning into galaxies.

What has become of the material in the forest clouds? Some may still be in the cool low-surface-density clouds detected in the vicinity of galaxies²⁶; some may fuel the rapid star formation observed in less luminous galaxies at redshifts $z < 1$, and the winds from supernovae will have redistributed some of this matter as hotter plasma; some may have settled onto the gaseous haloes of giant galaxies, contributing to systems like that pictured in Fig. 1; and some will have ended up in the plasma in groups and clusters of galaxies. One way or another, this material is now gathered together in pockets—concentrations of galaxies.

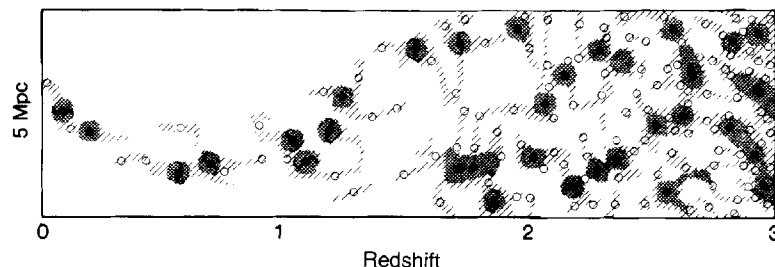
The forest clouds may be simple enough to allow analysis from first principles. Idealized models^{79–81} have included various geometries: spheres, slabs, filaments; states of motion ranging from quasistatic to expanding, collapsing and velocity caustics; and confinement by the pressure of hot intergalactic gas, or by the gravity of dark haloes, or simply by inertia. Recent numerical simulations suggest that the forest is indeed all of these things. Computations^{82–84} now have reached a level of detail that may be capable of taking account of all the relevant physical interactions, with the possibly serious exceptions of the feedback from star formation that we know is so important at the present epoch, and radiative transfer which is important in self-shielded, high-column systems optically thick to ionizing radiation. It is encouraging that the numerical simulations predict about the correct hydrogen column density distribution, coherence lengths and linewidth parameters for the absorption lines at redshifts $z \approx 2$ to $z \approx 5$, the largest observed for quasars. Figure 4 shows an example of the computed wispy gas distribution in the forest at $z = 2.5$.

In these models structure formation commenced not much earlier than $z = 5$. If feedback from star formation were an important factor in slowing evolution in the forest clouds it could mean substantial cosmic evolution commenced at still higher redshift.

Summary: cosmic evolution since $z = 3$

Perhaps the largest difference a traveller back in time to $z = 1$ would notice is the effect of the general expansion of the Universe: the mean number density of giant galaxies then was an order of magnitude larger than it is now. The traveller also would see that the less massive galaxies are brighter because stars are forming

FIG. 5 A cartoon sketch of cosmic evolution in a section of space of width and depth 5 Mpc. At redshift $z = 3$ matter has already been organized in clouds, the largest containing young elliptical galaxies, and the distribution of low-surface-density material has coherence length comparable to the distance between galaxies. By $z = 1$ the assembly of spiral disks is complete and much of the gas has been turned into stars. By the present epoch intergalactic space has been quite thoroughly cleared except near concentrations of galaxies. Further aspects of this sketch are discussed in the text.



faster. But if he travelled all the way back to $z = 3$ he would see a good deal more diffuse gas, much of it concentrated around star clusters recognizable as young galaxies, and the young galaxies would not appear strongly clustered because there is much less room for clustering than there is now.

We add some details to this picture in Fig. 5, which represents a section of space 5 Mpc by 5 Mpc on the narrow dimensions. A cube of this width at the present epoch contains on average about one giant galaxy (though there are large fluctuations around the mean). The length unit in the sketch is physical, so the mean number density of galaxies in a section this large varies as $(1+z)^3$ back to the epoch of galaxy assembly. The black circles represent ellipticals and the spheroid components in spirals. The evidence is that the formation of ellipticals as galaxies of stars was well under way at $z = 3$, and we have suggested that the similar-looking spheroid components in spirals formed early too. (To be visible the spheroids have to be drawn unrealistically large compared to the distance between galaxies; the same is true of the gaseous haloes of the galaxies.)

We indicate gas clouds with column densities $\Sigma_{\text{H I}} \geq 10^{17}$ atoms cm^{-2} by cross-hatched regions. At $z \lesssim 1$ these are found only around large galaxies^{21,22}; at higher redshifts they are the Lyman limit and damped Lyman- α clouds which we represent as being wrapped around young ellipticals and spheroids. Common giant spiral galaxies are represented by S-shaped curves centred on spheroids. Following the arguments presented above, we represent most of the disks seen today as having originated at redshifts between $z = 2$ and $z = 1$. It is likely that there were earlier generations of disks in the damped Lyman- α clouds that were disrupted by major accretion events.

The measured small-scale correlation of the galaxy positions is consistent with statistically stable clustering at $z \lesssim 1$ (ref. 85). Figure 5 therefore shows the nearest-neighbour distance about independent of time, and we have continued the trend back to $z = 3$. The result is that the voids between concentrations of galaxies are noticeably smaller at higher redshifts. The damped Lyman- α clouds at $z = 3$ cluster more strongly than expected from this simple picture for clustering evolution, and the young galaxies with spectra dominated by the Lyman- α emission line cluster very strongly around damped Lyman- α clouds⁸⁶. This may be a result of biasing: more massive protogalaxies tend to form in concentrations where the overall density is larger than average⁸⁷. Hence a more realistic picture might show stronger clustering of the high column density gas at $z = 3$. It is reasonable to suppose that the more massive spheroids that end up as elliptical galaxies tend to form in regions of higher ambient density, hence form in more strongly correlated positions, and so tend to end up in unusually dense regions, as observed⁸⁸.

The more widely spaced parallel lines in the sketch represent gas clouds at the column densities $\Sigma_{\text{H I}} \approx 10^{14} - 10^{17}$ atoms cm^{-2} characteristic of the forest clouds. There is still an appreciable, if much reduced, quantity of such gas at the present epoch. As we have attempted to represent in Fig. 5, this gas may be more than a megaparsec away from the nearest galaxy but it does tend to avoid the large-scale voids in the galaxy distribution²⁶. The general expansion of the Universe brings this arrangement of low-surface density gas clouds at the present epoch back to a froth that comes close to filling space at $z = 3$. This is a crude picture for the

evolution of the forest clouds. A more detailed picture would take account of the fact that some forest material is being accreted by damped Lyman- α clouds and converted to stars in giant galaxies, some is turning into dwarf galaxies, and galactic winds and accretion shocks are redistributing some as hot plasma.

We represent dwarf galaxies in Fig. 5 by the open circles. Because the forest clouds are probably feeding star formation in dwarf galaxies, we have placed the dwarfs in or near these clouds at high redshifts. A time traveller would notice that irregular galaxies are more common at $z = 1$. Some of the dwarfs that contribute to the large optical counts of galaxies will by now have fallen into larger galaxies, some will have been disrupted by mass loss driven by supernovae, and others, perhaps the majority, undoubtedly have survived as inconspicuous dwarfs.

At the present epoch the dwarf galaxies seem to avoid the voids defined by the giants as strongly as do gas clouds⁸⁹. The likely explanation is that gravity has drawn together galaxies and gas to produce the present clustering of matter, in the process emptying the voids. If this is so, then any diffuse plasma at temperatures less than 10^6 K would have been swept out of the voids too. This is why we do not have an entry for smooth void plasma in Table 1.

Theories of galaxy formation

Neither our discussion nor our synthesis in Figs 3 and 5 have made use of *ab initio* theories for the origin of the galaxies and their large-scale distribution—the familiar ‘package deals’ combining models for the dark matter, the nature of the primeval departures from exact homogeneity, and the background cosmology. That is in part because predictions of the theories are not yet very specific. For example, theories^{82,83} for the origin of the forest clouds yield about equally successful results if the total mass density parameter, baryonic plus nonbaryonic, is $\Omega = 1$ or $\Omega = 0.2$. The reason is that the parameters are tuned to produce the observed mass distribution on the scales of the galaxies; the simulations suggest that the Lyman- α forest forms as a natural accompaniment to galaxy formation out of primeval gas in any scenario.

A viable *ab initio* theory must be able to account for the observations, of course, and here are three examples of the challenges it must meet. First, the dramatic clearing of the voids between redshift $z = 3$ and the present is a demanding constraint, and may be a particular problem for theories in which the total mass density in baryons plus cold dark matter is $\Omega = 1$. It is sometimes argued that dynamical measures (based on the relative motions of galaxies) underestimate the mean mass density because most of the mass is in intergalactic space and does not affect the relative accelerations of neighbouring galaxies referred to the general expansion. But if intergalactic space today were full of invisible haloes, why are they not seen today accreting absorption clouds as they do at high redshift: why is intergalactic space so bare of baryon clouds and dwarf galaxies? The complementary problem if $\Omega = 1$ is that clusters of galaxies seem to have more than their share of baryons per dark matter compared to the predicted cosmic mean²⁹. Second, if we add enough hot dark matter to avoid unreasonably high concentrations of mass around galaxies (as in the ‘mixed’ dark matter model), full-sized galaxies may not be assembled until $z \lesssim 1$ (refs 90, 91). This may agree with the rapid evolution of the intergalactic medium and the damped Lyman- α clouds at $z \lesssim 2$, but not with the evidence that many

giant galaxies look quite mature at $z = 1$, with very little building material left over in intergalactic space⁹², and it does not agree with the properties of the damped Lyman- α clouds or what look like many well-developed, if young, giant elliptical galaxies at $z = 3$. Third, the computations indicate the forest clouds evolve rapidly (faster than the Hubble rate) once collapse commences. That means that models designed to assemble galaxies at high redshift^{93,94} may agree with the early assembly of ellipticals but are in danger of producing and dissipating the forest clouds too early. As

these examples show, the observations are leading to some sharp tests of models for the origin of the galaxies, and might even be expected to be beneficial guides to the builders of new *ab initio* theories. $\square$

M. Fukugita is at the Institute for Advanced Study, Princeton, New Jersey 08540, USA and at the Yukawa Institute of Kyoto University, Kyoto 606, Japan. C. J. Hogan is at the University of Washington, Seattle, Washington 98195, USA. P. J. E. Peebles is at Princeton University, Princeton, New Jersey 08544, USA.

1. Peebles, P. J. E., Schramm, D. N., Turner, E. L. & Kron, R. G. *Nature* **352**, 769–776 (1991).
2. Kennicutt, R. C. *Astrophys. J.* **388**, 310–327 (1992).
3. Leitherer, C. & Heckman, T. M. *Astrophys. J. Suppl. Ser.* **96**, 9–38 (1995).
4. Griffiths, R. E. *et al. Astrophys. J.* **435**, L19–L22 (1994).
5. Glazebrook, K., Ellis, R., Santiago, B. & Griffiths, R. R. *Mon. Not. R. astr. Soc.* **275**, L19–L22 (1995).
6. Driver, S. P. *et al. Astrophys. J.* **449**, L23–L27 (1995).
7. Crampton, D., Le Fevre, O., Lilly, S. J. & Hammer, F. *Astrophys. J.* **455**, 96–107 (1995).
8. Lilly, S. J. *et al. Astrophys. J.* **445**, 108–124 (1995).
9. Djorgovski, S. *et al. Astrophys. J.* **438**, L13–L16 (1995).
10. Glazebrook, K., Peacock, J. A., Miller, L. & Collins, C. A. *Mon. Not. R. astr. Soc.* **275**, 169–184 (1995).
11. Gardner, J. P., Cowie, L. L. & Wainscoat, R. J. *Astrophys. J.* **415**, L9–L12 (1993).
12. Steidel, C. C., Dickinson, M. & Persson, S. E. *Astrophys. J.* **437**, L75–L78 (1994).
13. Hernquist, L. & Mihos, J. C. *Astrophys. J.* **448**, 41–63 (1995).
14. Yun, M. S., Ho, P. T. P. & Lo, K. Y. *Nature* **372**, 530–532 (1994).
15. Butcher, H. & Oemler, A. Jr *Astrophys. J.* **285**, 426–438 (1984).
16. Koo, D. C. & Kron, R. G. *A. Rev. Astr. Astrophys.* **30**, 613–652 (1992).
17. Broadhurst, T. J., Ellis, R. S. & Shanks, T. *Mon. Not. R. astr. Soc.* **235**, 827–856 (1988).
18. Glazebrook, K. *et al. Mon. Not. R. astr. Soc.* **273**, 157–168 (1995).
19. Gronwall, C. & Koo, D. C. *Astrophys. J.* **440**, L1–L4 (1995).
20. Unavane, M., Wyse, R. G. F. & Gilmore, G. *Mon. Not. R. astr. Soc.* **278**, 727–736 (1996).
21. Bergeron, J. & Boissé, P. *Astr. Astrophys.* **243**, 344–366 (1991).
22. Steidel, C. C. in *The Environment and Evolution of Galaxies* (eds Shull, J. M. & Thronson, H. A. Jr) 263 (Kluwer, Dordrecht, 1993).
23. Lanzetta, K. M., Bowen, D. V., Tytler, D. & Webb, J. K. *Astrophys. J.* **442**, 538–568 (1995).
24. Morris, S. L. *et al. Astrophys. J.* **419**, 524–540 (1993).
25. Salpeter, E. E. & Hoffman, G. L. *Astrophys. J.* **441**, 51–57 (1995).
26. Stocke, J. T., Shull, J. M., Penton, S., Donahue, M. & Canilli, C. *Astrophys. J.* **451**, 24–43 (1995).
27. Rao, S. & Briggs, F. *Astrophys. J.* **419**, 515–523 (1993).
28. Fall, S. M. & Pei, Y. C. *Astrophys. J.* **402**, 479–492 (1993).
29. White, S. D. M., Navarro, J. F., Evrard, A. E. & Frenk, C. S. *Nature* **366**, 429–433 (1993).
30. White, D. A. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **273**, 72–84 (1995).
31. Pildis, R. A., Bregman, J. N. & Evrard, A. E. *Astrophys. J.* **443**, 514–526 (1995).
32. Mulchaey, J. S., Davis, D. S., Mushotzky, R. F. & Burstein, D. *Astrophys. J.* **456**, 80–97 (1996).
33. Salucci, P. & Persic, M. *Mon. Not. R. astr. Soc.* (in the press).
34. Alcock, C. *et al. Phys. Rev. Lett.* **74**, 2867–2870 (1995).
35. Copi, C. J., Schramm, D. N. & Turner, M. S. *Science* **267**, 192–199 (1995).
36. Songaila, A., Cowie, L. L., Hogan, C. J. & Rugers, M. *Nature* **368**, 599–604 (1994).
37. Carswell, R. F., Rauch, M., Weymann, R. J., Cooke, A. J. & Webb, J. K. *Mon. Not. R. astr. Soc.* **268**, L1–L4 (1994).
38. Rugers, M. & Hogan, C. J. *Astrophys. J.* **459**, L1–L5 (1996).
39. McCarthy, P. J. *A. Rev. Astr. Astrophys.* **31**, 639–688 (1993).
40. Lacy, M. *et al. Mon. Not. R. astr. Soc.* **271**, 504–512 (1994).
41. Spinrad, H., Dey, A. & Graham, J. R. *Astrophys. J.* **438**, L51–L54 (1995).
42. Daly, R. A. *Astrophys. J.* **399**, 426–443 (1992).
43. Dunlop, J. *et al. Nature* (in the press).
44. Cowie, L. L., Hu, E. & Songaila, A. *Nature* **377**, 603–605 (1995).
45. Steidel, C. C., Giallisco, M., Pettini, M., Dickinson, M. & Adelberger, K. L. *Astrophys. J.* **462**, L17–L21 (1996).
46. Wolfe, A. M. *Ann. N.Y. Acad. Sci.* **688**, 281–296 (1993).
47. Wolfe, A. M. *et al. Astrophys. J.* **435**, L101–L104 (1994).
48. Lu, L., Sargent, W. L. W., Womble, D. S. & Barlow, T. A. *Astrophys. J.* **457**, L1–L4 (1996).
49. Crotts, A. P. S., Bechtold, J., Fang, Y. & Duncan, R. C. *Astrophys. J.* **437**, L79–L82 (1994).
50. Lanzetta, K. M., Wolfe, A. M. & Turnshek, D. A. *Astrophys. J.* **440**, 435–457 (1995).
51. Wolfe, A. M., Lanzetta, K. M., Foltz, C. B. & Chaffee, F. H. *Astrophys. J.* **454**, 698–725 (1995).
52. Pei, Y. C. & Fall, S. M. *Astrophys. J.* **454**, 69–76 (1995).
53. Elston, R., Bechtold, J., Hill, G. J. & Ge, J. *Astrophys. J.* **456**, L13–L16 (1996).
54. Pettini, M., Lipman, K. & Hunstead, R. W. *Astrophys. J.* **451**, 100–110 (1995).
55. Lu, L., Savage, B. D., Tripp, T. M. & Meyer, D. M. *Astrophys. J.* **447**, 597–615 (1995).
56. Pettini, M., Smith, L. J., Hunstead, R. W. & King, D. L. *Astrophys. J.* **426**, 79–96 (1994).
57. Ostriker, J. P. & Thuan, T. X. *Astrophys. J.* **202**, 353–364 (1975).
58. Gilmore, G., Kuijken, K. & Wyse, R. F. G. *A. Rev. Astr. Astrophys.* **27**, 555–627 (1989).
59. Majewski, S. R. *A. Rev. Astr. Astrophys.* **31**, 575–638 (1993).
60. Steidel, C. C., Pettini, M. & Hamilton, D. *Astr. J.* **110**, 2519–2536 (1995).
61. Foltz, C. B., Hewett, P. C., Chaffee, F. H. & Hogan, C. J. *Astr. J.* **105**, 22–29 (1993).
62. Francis, P. J. & Hewett, P. C. *Astr. J.* **105**, 1633–1636 (1993).
63. Dinshaw, N. & Impey, C. D. *Astr. J.* (submitted).
64. Wolfe, A. M. *Astrophys. J.* **402**, 411–419 (1993).
65. Fall, S. M., Cheriot, S. & Pei, Y. C. *Astrophys. J.* (in the press).
66. Schneider, D., Schmidt, M. & Gunn, J. E. *Astr. J.* **102**, 837–840 (1991).
67. Hewett, P. C. & Foltz, C. B. *Publ. astr. Soc. Pacif.* **106**, 113–130 (1994).
68. Warren, S. J., Hewett, P. C. & Osmer, P. S. *Astrophys. J.* **421**, 412–433 (1994).
69. Sargent, W. L. W., Young, P. J., Bokserberg, A. & Tytler, D. *Astrophys. J. Suppl. Ser.* **42**, 41–81 (1980).
70. Songaila, A. & Cowie, L. L. *Astr. J.* (submitted).
71. Madau, P. & Meiksin, A. *Astrophys. J.* **433**, L53–L56 (1994).
72. Jakobsen, P. *et al. Nature* **370**, 35–39 (1994).
73. Davidsen, A. F., Kriss, G. A. & Zheng, W. *Nature* **380**, 47–49 (1996).
74. Smette, A. *et al. Astrophys. J.* **389**, 39–62 (1992).
75. Dinshaw, N., Foltz, C. B., Impey, C. D., Weymann, R. J. & Morris, S. L. *Nature* **373**, 223–225 (1995).
76. Dinshaw, N., Impey, C. D., Foltz, C. B., Weymann, R. J. & Chaffee, F. H. *Astrophys. J.* **437**, L87–L90 (1994).
77. Bechtold, J., Crotts, A. P. S., Duncan, R. C. & Fang, Y. *Astrophys. J.* **437**, L83–L86 (1995).
78. Cowie, L. L., Songaila, A., Kim, T.-S. & Hu, E. M. *Astr. J.* **108**, 1522–1530 (1995).
79. Miralda-Escudé, J. & Rees, M. J. *Mon. Not. R. astr. Soc.* **266**, 343–352 (1994).
80. Charlton, J. C., Salpeter, E. E. & Linder, S. M. *Astrophys. J.* **430**, L29–L32 (1994).
81. Mo, H. J., Miralda-Escudé, J. & Rees, M. J. *Mon. Not. R. astr. Soc.* **264**, 705–711 (1993).
82. Cen, R., Miralda-Escudé, J., Ostriker, J. P. & Rauch, M. *Astrophys. J.* **437**, L9–L12 (1994).
83. Hernquist, L., Katz, N., Weinberg, D. H. & Miralda-Escudé, J. *Astrophys. J.* **457**, L51–L55 (1996).
84. Katz, N., Weinberg, D. H., Miralda-Escudé, J. & Hernquist, L. *Astrophys. J.* **457**, L57–L60 (1996).
85. le Fevre, O. *et al. Archive preprint astro-ph 9510090* (1995).
86. Wolfe, A. M. *Astrophys. J.* **492**, 411–419 (1993).
87. Kaiser, N. *Astrophys. J.* **284**, L9–L12 (1985).
88. Dressler, A. *Astrophys. J.* **236**, 351–365 (1980).
89. Peebles, P. J. E. *Principles of Physical Cosmology* Sect. 3 (Princeton Univ. Press, 1993).
90. Klypin, A., Borgani, S., Holtzman, J. & Primack, J. *Astrophys. J.* **444**, 1–14 (1995).
91. Cen, R. & Ostriker, J. P. *Astrophys. J.* **431**, 451–476 (1994).
92. Ma, C. P. & Bertschinger, E. *Astrophys. J.* **434**, L5–L9 (1994).
93. Peebles, P. J. E. *Astrophys. J.* **432**, L1–L4 (1994).
94. Hogan, C. J. *Phys. Rev. Lett.* **74**, 3105–3108 (1995).

ACKNOWLEDGEMENTS. This Review Article was inspired by the 1995 Aspen Winter Conference on Astrophysics; we thank the participants for their guidance. We are grateful to M. Dickinson, S. M. Fall, N. Katz, K. Y. Lo, C. KcKee, J. Ostriker, J. M. Shull, J. Silk, C. Steidel and A. Wolfe for comments and for help with the figures. The work was supported in part at the University of Washington by NASA and the US NSF; and at Princeton University by the US NSF. M.F. thanks the Fuji Xerox Corporation for generous support.

Guide to authors of contributions to *Nature*

Nature is a weekly international journal covering all the sciences. It is intended for an interdisciplinary readership, so all manuscripts should be written clearly and simply. Contributors should pay particular attention to the fact that English is not the first language of many readers. Space is limited and competition severe, so brevity is highly valued.

Nature publishes the following types of article:

Review articles survey recent developments in a topical area of scientific research. Contributors wishing to submit a review should first send a one-page synopsis to the Reviews Coordinator.

- Reviews inform a broad readership about fields where there are rapid, important advances.

- They focus on one topical aspect of a field rather than providing a comprehensive literature survey.

- They can be controversial but in this event should briefly indicate opposing viewpoints. They should not be focused on the author's own work.

- Language should be simple, novel concepts defined and jargon explained.

- Reviews should not be more than 6,000 words and ideally should be shorter. There should be no more than 100 references and ideally half that number. There is no limit to the number of display items or explanatory boxes (used for clarification of technical points or for background material), but reviews do not generally take more than six pages of *Nature*.

- Review articles are often substantially edited by *Nature*'s editors in consultation with the author.

Progress articles review particularly topical and fast-moving fields for a nonspecialist readership. They are similar to Reviews except for the following:

- They do not exceed four pages of *Nature* in length, including display items and references. (One printed page of text is about 1,300 words.)

- They focus on current papers of outstanding interest that are setting new standards in a field.

- Because of their topicality, Progress articles should be written and submitted within a few weeks of *Nature*'s editors expressing an interest in a synopsis.

- Authors may discuss their own work, but should make it clear in the text if they are presenting a personal, rather than a consensus, view.

- Titles are brief (generally a single line) and relatively informal.

Articles are reports of original scientific research.

- They report novel conclusions of broad general scientific interest.

- They represent a substantial advance in scientific understanding of an important problem.

- They should not be longer than 3,000 words, have more than six display items (with captions of fewer than 300 words) or have more than 50 references.

- They have a 'heading' of up to 80 words, often rewritten by *Nature*'s editors, which advertises the content of the paper in general terms. The heading should not contain numbers, abbreviations or measurements unless central to the message of the paper.

- The introduction and summary should be contained in the first two or three paragraphs of the main text.

Letters to *Nature* are short reports focused on a novel, outstanding finding.

- The implications of the main result are of interest to nonspecialists.

- They should not exceed 2 1/2 pages of *Nature*, equivalent to 1,500 words of text with four small display items that have brief captions.

- They have a bold-text first paragraph of not more than 150 words summarizing the rationale for the study, the main result and conclusion. This paragraph is aimed particularly at nonspecialist readers.

- Letters should have fewer than 30 references and no more than four display items. Captions should not exceed 300 words.

- **Articles and Letters** contain a statement at the end of the text: "Correspondence and requests for materials to xxxxx". Database accession numbers should be included after this statement.

Preparation of manuscripts

- All manuscripts should be typed, double-spaced, on one side of the paper only.

- Manuscripts should be accompanied by a brief cover letter from the corresponding author, containing a full postal address, telephone and fax number, and e-mail address.

- Five copies of manuscripts and original figures are needed, together with two copies of the covering letter.

- Five copies of relevant related manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such.

- Unless otherwise instructed by the editor handling a manuscript, when sending revised or resubmitted manuscripts, five copies are required, each accompanied by a copy of the authors' response to ref-

erees' and editors' comments on the earlier version.

- Titles should be brief, pertinent and simple, avoiding active verbs, numerical values, abbreviations and punctuation.

- References are numbered sequentially as they appear in the text, tables and figure legends. Only papers that are published or in press should be given numbers: manuscripts submitted or in preparation should be mentioned in the text with a list of authors.

- Reference lists contain only citations to published papers, and do not contain textual material, grant details or acknowledgements.

- Acknowledgements must be brief and appear after the reference list. *Nature* does not publish grant contribution numbers.

- Figures should not be larger than 22 by 28 cm unless unavoidable. They should be marked with the author's name and, where known, the manuscript reference number. One photocopy of the original figures should be provided. All original figures are returned when *Nature* cannot offer to publish a manuscript, but one copy of the manuscript and the photocopy of the figures are kept in confidential files for three months and then destroyed. Detailed instructions for preparation of figures are available from the production department, and digital files of final versions of the figures may be requested.

- Colour figures are welcome but a contribution towards the cost of reproduction is requested. Inability to pay this charge will not prevent *Nature* publishing figures where colour is essential.

- Figures containing protein/nucleotide sequence information should ideally use the three-letter code for amino acids. One column width can accommodate 20 amino acids or 60 base pairs.

- Figure legends should contain fewer than 300 words. They should consist of a brief description of the figure (title, explanation of the parts, explanation of symbols) followed by a telegraphic account of the methods, if appropriate. Multipart figures are discouraged unless the parts are logically connected.

- Tables do not have a 'methods' section. Symbols and abbreviations in the table should be defined immediately below the table, followed by essential descriptive material, all in double-spaced text. Tables should each be presented on a separate sheet of paper.

- As a condition of publication, authors are required to make materials and methods used freely available to academic researchers for their own use. Supporting datasets must be made available at the time of publication either by deposition in the appropriate public database or by distribution on the Internet, together with the relevant accession numbers or site address. In the case of X-ray crystallographic coordinates only, public access may be delayed for up to one year after publication.

- Manuscripts can be submitted to the Editor at *Nature*, Porter's South, Crinan St, London N1 9XW or at *Nature*, 968 National Press Building, Washington, DC 20045-1938. Proofs should be returned by express mail to London.

Other contributions

Nature publishes informal material in each issue, as follows.

- **Correspondence** contributions are short comments on topical issues, anecdotal material or reactions of readers to nontechnical material published in *Nature* (mainly News, Opinion, Book Review or Commentary articles). All are unsolicited; submissions by e-mail to nature@nature.com are preferred (accents on names and places should be clearly indicated). Contributions should be brief (fewer than 500 words).

- **Commentary** articles are informal, editorial comments on topical issues of public interest that have some bearing on scientific research. They are often controversial. Unsolicited manuscripts are considered, but enquiries can be made to the Commentary Editor by e-mail (nature@nature.com) or fax before formal submission as a one-page synopsis or informal letter.

- **News and Views** editorials inform nonspecialist readers about new scientific advances, either as reported in published papers or in the form of meeting reports. Most are commissioned, but proposals can be made to the News and Views Editor in advance of publication of the paper or of the meeting concerned. Authors are not allowed to discuss their own work or work from elsewhere in their laboratory. Detailed guidelines are available on request.

- **Scientific Correspondence** is for informal discussion of topical scientific issues, including material published in *Nature*. Priority is given to contributions of fewer than 500 words. Detailed guidelines are available on request. All contributions are unsolicited.

- **Book Reviews** are all commissioned by the Book Review editor. Unsolicited contributions are not considered.

- *Nature*'s editors will request final copy for Commentary, News and Views, Scientific Correspondence and Book Reviews on diskette or via e-mail.

A more detailed Guide to Authors is available on *Nature*'s www site:
<http://www.nature.com>
<http://www.america.nature.com>

Very low braking index for the Vela pulsar

A. G. Lyne, R. S. Pritchard, F. Graham-Smith & F. Camilo

The University of Manchester, NRAL, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

THE rotation rate of a pulsar is thought to decrease with time according to a simple power law, with a 'braking index' equal to 3 if rotational energy is lost through radiation from a dipolar magnetic field¹⁻³. The age of the pulsar can accordingly be determined simply by measuring the current rotation rate, and its current rate of change. Here we report an analysis of the rotation rate of the Vela pulsar as observed over 25 years. We find that the braking index is 1.4 ± 0.2 , suggesting that the braking cannot be attributed entirely to radiation from a constant magnetic dipole but is probably due to a changing, magnetic moment or effective moment of inertia. Taken at face value, the result implies that the Vela pulsar may be much older than previously thought, and that inferred velocities of the supernova ejecta⁴ and an X-ray jet from the pulsar⁵ are correspondingly reduced. If other young pulsars associated with supernova remnants have similarly low braking indices, then they too may be much older than believed, thereby reducing their estimated velocities.

Pulsars are assumed to spin down according to the relation $\dot{\nu} = -K\nu^n$, where ν is the rotation rate, n is the braking index and K depends on the magnetic moment and moment of inertia of the neutron star. On the assumption that it had a very high initial rotation rate, the age of the pulsar is $\tau = -\nu/(n-1)\dot{\nu} = -\nu/2\dot{\nu}$. Assuming K is a constant, differentiation of the slow-down power law gives $n = \nu\ddot{\nu}/(\dot{\nu})^2$, so that in principle n can be obtained from a measurement of $\ddot{\nu}$. But such a measurement has only been possible hitherto for the three youngest pulsars; the Crab (PSR B0531 + 21), PSR B0540 - 69 and PSR B1509 - 58, giving values for n of 2.51 ± 0.01 , 2.24 ± 0.04 and 2.837 ± 0.001 , respectively⁶⁻⁸. For the Vela pulsar and older pulsars, this has not been possible because of the high levels of glitch activity (see below) and timing noise which dominate the steady evolution of the rotation.

The Vela pulsar, with $\nu \approx 11.2$ Hz and $\dot{\nu} \approx -157 \times 10^{-13}$ Hz s⁻¹, suffered nine major glitches between 1969 and 1994, each characterized observationally by a sudden increase in rotation rate $\Delta\nu/\nu \approx 2 \times 10^{-6}$ and in slow-down rate $\Delta\dot{\nu}/\dot{\nu} \gtrsim 2 \times 10^{-2}$ and followed by both short-term quasi-exponential transients and long-term relaxation. The glitches can be seen in Fig. 1 which shows the slow-down rate $\dot{\nu}$ throughout 25 years. The presence of the transients is usually taken to indicate the existence of a superfluid neutron component within the neutron star^{9,10}. The glitches themselves are thought to be caused by the erratic transfer of angular momentum, in the form of quantized vortices, to the solid crust of the star whose rotation we observe⁹.

The short-term transients have been resolved in the more recent glitches into three exponentially decaying components with time constants of the order of 0.4, 3.2 and 32 days (refs 10-12). These transients are variable in size and are not equally resolved at each glitch. Over longer timescales the relaxation continues, with the magnitude of the slowdown rate between glitches decreasing monotonically but at a rate which varies slowly, possibly due to the effects of timing noise or micro-glitches¹¹. But it will be observed that, although the slopes and amplitudes of the saw-tooth waveform in Fig. 1 vary, the slowdown rate returns to essentially the same level immediately after the short-term decay following a glitch. For instance, the values of $\dot{\nu}$ 150 days after each glitch are indicated in the figure by the small filled symbols; this interval is long enough for most of the large-

amplitude short-term transient terms to subside, but short enough that contamination by timing noise is minimized.

A least-squares straight-line fit to this series of values of $\dot{\nu}$ yields a value of $\dot{\nu} = (35 \pm 7) \times 10^{-24}$ Hz s⁻² and $n = 1.6 \pm 0.3$. This result is robust, in that analyses based on values of $\dot{\nu}$ determined 50, 100, 200 and 250 days after each glitch provide values of $n = 1.7 \pm 0.6$, 1.5 ± 0.4 , 1.5 ± 0.4 and 1.5 ± 0.5 , that is, the same result within the (larger) uncertainties. But an even more precise result, $\dot{\nu} = (31 \pm 4) \times 10^{-24}$ Hz s⁻² and $n = 1.4 \pm 0.2$, is obtained by extrapolating the values of $\dot{\nu}$ 150 days after each glitch back to the corresponding glitch epochs. We do this using local mean inter-glitch values of $\dot{\nu}$ obtained from a fit to data collected between 150 days after a glitch and the succeeding glitch, employing this procedure to minimize the biases in $\dot{\nu}$ introduced by the long-term relaxation observable in Fig. 1. The value of n determined from the values of $\dot{\nu}$ immediately before each glitch is 1.9 ± 1.2 , the large uncertainty arising from the wide scatter in pre-glitch $\dot{\nu}$ values. We can understand this large scatter as arising from the combined effect of large inter-glitch slopes in $\dot{\nu}$ and variable time intervals since the preceding glitches, together with timing noise. In summary, it seems that the underlying slow-down rate is recovered at the epoch of each glitch, and the subsequent short-term transients are then merely the response of a loosely coupled fluid stellar component.

Integration of the power-law slow-down equation, assuming constant K , gives the age of the pulsar:

$$\tau = -\frac{\nu}{(n-1)\dot{\nu}} \left[1 - \left(\frac{\nu}{\nu_i} \right)^{n-1} \right] \quad (1)$$

If $\nu \ll \nu_i$, the rotation rate at birth, equation (1) reduces to $\tau = -\nu/(n-1)\dot{\nu}$ if $n > 1$, and to the characteristic age $\tau_c = -\nu/2\dot{\nu}$ if $n = 3$, usually regarded as an upper limit to the age of a pulsar¹³. Our measurement of n for Vela removes one assumption from the age calculation ($n = 3$) and leads to $-\nu/(n-1)\dot{\nu} > 38$ kyr for $1.2 < n < 1.6$. However, the initial spin period $P_i = 1/\nu_i$ is unknown. Assuming $P_i = 20$ ms, the estimated initial period of the Crab pulsar¹³, the age derived for Vela is 22-29 kyr for $1.6 > n > 1.2$, which is 2-3 times the usual

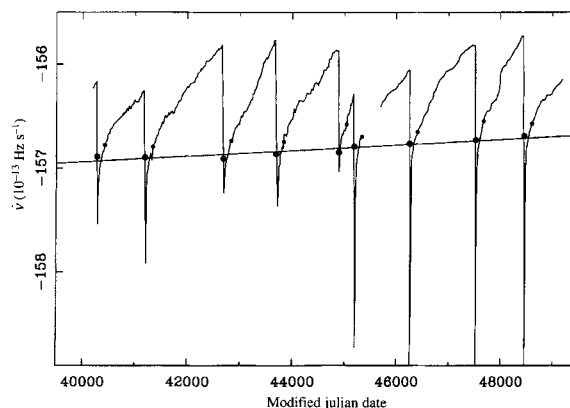


FIG. 1 The slowdown rate $\dot{\nu}$ of the Vela pulsar over a 25-year period. Data for this figure were obtained from local fits for ν and $\dot{\nu}$ to the arrival times published by Downs and Reichley²² and Downs and Krause-Polstorff²³, and the values of $\dot{\nu}$ published by Flanagan²⁴. Each glitch is followed by a short-term transient relaxation, followed by a longer-term decrease in magnitude of the slowdown rate. The values of $\dot{\nu}$ 150 days after the nominal epoch of each of nine glitches are shown by small filled symbols. Large filled symbols show values of $\dot{\nu}$ extrapolated from these points back to the epoch of each glitch (see text). The best straight-line fit to these data is plotted, revealing an unexpectedly small but significant positive slope, giving $\dot{\nu} = (31 \pm 4) \times 10^{-24}$ Hz s⁻².

upper limit of $\tau_c = 11$ kyr calculated for $P_i = 0$ and $n = 3$. The age is greater still if $P_i < 20$ ms. Some recent estimates of the ages of the Vela supernova remnant and pulsar are based on studies identifying the site of the supernova explosion¹⁴ together with the position offset and proper motion¹⁵ of the pulsar, yielding values¹⁴ of 18–31 kyr. Our estimate of the Vela pulsar age is therefore broadly consistent with independent age determinations, if the initial pulsar spin period was not much larger than one-third of its present value, and lends support to the conclusion that the true age of the Vela system is higher than the maximum of 11 kyr previously supposed by a factor of 2–3.

The increase in the estimated age of the Vela pulsar and supernova remnant by such a factor has several important consequences for many studies of the dynamics and evolution of the remnant. For example, the “bullets” ejected by the supernova⁴ are now moving slower by the same factor, with consequences for their interpretation and inferences about the ambient medium. The X-ray jet detected from the Vela pulsar, and suggested to account for ~50% of its spindown energy⁵, now carries much less energy: a characteristic velocity of the jet, v_b (proportional to the inverse age), factors in the determination of the jet power as v_b^3 , so that the jet power is reduced by about one order of magnitude.

We now consider possible physical causes of the very low braking index measured for the Vela pulsar. Assuming constant K in the spin-down law, a value of $n < 3$ may be explained qualitatively either by an outflow of charged particles that remove mechanical angular momentum from the star, or by strong magnetospheric currents that distort the otherwise dipolar magnetic field¹⁶. Whether such effects can plausibly have the magnitude required to give $n = 1.4 \pm 0.2$ is unclear. Alternatively, K is not constant, that is, the correct spin-down law is of the form $\dot{v} = -K(t)v^n$. In this case the observed braking index, $n_{\text{obs}} = v\ddot{v}/(\dot{v})^2$, is not equal to the braking index n of the spin-down law; instead¹⁶

$$n_{\text{obs}} = n + \frac{v\dot{K}(t)}{\dot{v}K(t)} = n - (7.1 \times 10^{11}) \frac{\dot{K}(t)}{K(t)} \quad (2)$$

With an assumed value for n (for example, 3), equation (2) gives the timescale for the variation, $K/\dot{K}$, about 14 kyr for Vela. From equation (2) and our measurement $n_{\text{obs}}, \dot{K} > 0$ if $n > 1.6$. Integration of the spin-down law in this case requires the functional form of $K(t)$ so that an age for the pulsar cannot be estimated from its spin parameters without further assumptions. If, however, $\dot{K} > 0$ since the birth of the pulsar, as it may be today, the age estimates calculated following equation (1) are only lower limits. For magnetic braking $K \propto M^2/I$, so the increase in K may be due to an increasing magnetic moment M or decreasing effective moment of inertia I . In the Crab pulsar, there is a permanent increase in the spin-down rate at each glitch⁶ by an amount

$\Delta\dot{v}/\dot{v} \approx 3 \times 10^{-4}$, suggesting that the spindown torque acts on less moment of inertia after such glitches¹⁷ and implying a progressive decrease in I . Similar persistent shifts in $\Delta\dot{v}$ occurring in Vela would be undetectable on the scale of Fig. 1 and it is possible that the low value of n_{obs} is in fact caused by a decreasing I . This could occur, for instance, if some of the superfluid vortices became permanently pinned to the nuclei in the lattice of the solid crust of the star, so that they no longer participate in the slowdown¹⁷.

We note that whereas the three pulsars with a previous determination of n_{obs} have characteristic ages of 1–2 kyr and n_{obs} of 2.2–2.8 (refs 6–8), we have now determined that the Vela pulsar, with an age about 20 times greater, has $n_{\text{obs}} = 1.4 \pm 0.2$. If this broad correlation is not coincidental, it suggests that pulsars initially spin down with a value of n_{obs} somewhat below 3, acquiring even lower braking indices as they age. Another difference between the youngest pulsars and those with ages of 10–300 kyr is that the latter, like Vela, have much higher “glitch activity” than the former¹⁸, suggesting that glitch activity and small n_{obs} may be related. What their braking indices settle down to after pulsars age beyond about 10^6 yr remains a very important but speculative question.

Unless Vela is in some sense special, there is no reason why other youthful pulsars should not behave similarly, and we briefly consider some consequences. Lately, several pulsar-supernova remnant associations have been postulated; in a recent comprehensive study¹⁹, 15 of the 18 pulsars with the smallest values of τ_c have what is plausibly an associated supernova remnant. In some cases the association requires extremely large transverse pulsar velocities, as inferred from the angular displacement of the pulsar from the remnant centre and the assumed pulsar age τ_c . After disregarding a few questionable associations, Frail *et al.*¹⁹ obtain a mean transverse velocity for their sample of $V_t = 550 \text{ km s}^{-1}$. Furthermore, in some otherwise compelling associations, the velocity required is much higher, such as in the case of PSR B1757–24 and G5.4–1.2, with $V_t \approx 1,800 \text{ km s}^{-1}$ (ref. 20). Such high velocities are hard to understand and cause doubts on a few of the putative associations. If some of these pulsars, like Vela, are in fact older than previously thought owing to relatively small braking indices, the velocities inferred would be correspondingly smaller, helping to strengthen the case for some of the associations, and reducing the mean inferred transverse velocity of this sample of young pulsars. Somewhat lower velocities would also agree better with the recent demonstration²¹ that the mean transverse velocity of pulsars at birth is $V_t = 345 \pm 70 \text{ km s}^{-1}$. In the future, actual measurement of the velocities in some of the young systems, only feasible through interferometric techniques, may provide estimates of their ages and hence yield some information about braking indices. □

Received 6 December 1995; accepted 2 April 1996.

1. Pacini, F. *Nature* **219**, 145–146 (1968).
2. Ostriker, J. P. & Gunn, J. E. *Astrophys. J.* **157**, 1395–1417 (1969).
3. Goldreich, P. & Julian, W. H. *Astrophys. J.* **157**, 869–880 (1969).
4. Strom, R., Johnston, H. M., Verbunt, F. & Aschenbach, B. *Nature* **373**, 590–592 (1995).
5. Markwardt, C. B. & Ögelman, H. *Nature* **375**, 40–42 (1995).
6. Lyne, A. G., Pritchard, R. S. & Smith, F. G. *Mon. Not. R. astr. Soc.* **265**, 1003–1012 (1993).
7. Boyd, P. T. *et al.* *Astrophys. J.* **448**, 365–368 (1995).
8. Kaspi, V. M., Manchester, R. N., Siegelman, B., Johnston, S. & Lyne, A. G. *Astrophys. J.* **422**, L83–L86 (1994).
9. Baym, G., Pethick, C., Pines, D. & Ruderman, M. *Nature* **224**, 872–874 (1969).
10. Alpar, M. A., Chau, H. F., Cheng, K. S. & Pines, D. *Astrophys. J.* **409**, 345–359 (1993).
11. Cordes, J. M., Downs, G. S. & Krause-Polstorff, J. *Astrophys. J.* **330**, 847–869 (1988).
12. Chau, H. F., McCulloch, P. M., Nandkumar, R. & Pines, D. *Astrophys. J.* **413**, L113–L116 (1993).
13. Manchester, R. N. & Taylor, J. H. *Pulsar* (Freeman, San Francisco, 1977).

14. Aschenbach, B., Egger, R. & Trümper, J. *Nature* **373**, 587–590 (1995).
15. Bailes, M., Manchester, R. N., Kesteven, M. J., Norris, R. P. & Reynolds, J. E. *Astrophys. J.* **343**, L53–L55 (1989).
16. Blandford, R. D. & Romani, R. W. *Mon. Not. R. astr. Soc.* **234**, 57P–60P (1988).
17. Alpar, M. A. & Pines, D. in *Isolated Pulsars* (eds Van Riper, K. A., Epstein, R. & Ho, C.) 17–27 (Cambridge Univ. Press, 1993).
18. Lyne, A. G. in *The Lives of the Neutron Stars* (eds Alpar, A., Kızıloğlu, Ü. & van Paradis, J.) 167–176 (NATO ASI Ser., Kluwer, Dordrecht, 1995).
19. Frail, D. A., Goss, W. M. & Whiteoak, J. B. *Z. Astrophys. J.* **437**, 781–793 (1994).
20. Frail, D. A. & Kulkarni, S. R. *Nature* **352**, 785–787 (1991).
21. Lyne, A. G. & Lorimer, D. R. *Nature* **369**, 127–129 (1994).
22. Downs, G. S. & Reichley, P. E. *Astrophys. J. Suppl. Ser.* **53**, 169–240 (1983).
23. Downs, G. S. & Krause-Polstorff, J. *Astrophys. J. Suppl. Ser.* **62**, 81–107 (1986).
24. Flanagan, C. S. in *The Lives of Neutron Stars* (eds Alpar, A., Kızıloğlu, Ü. & van Paradis, J.) 181–184 (NATO ASI Ser., Kluwer, Dordrecht, 1995).

ACKNOWLEDGEMENTS. We thank D. Pines and A. Alpar for helpful suggestions about the interpretations presented here.

Synthesis of layered LiMnO_2 as an electrode for rechargeable lithium batteries

A. Robert Armstrong & Peter G. Bruce

School of Chemistry, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, UK

RECHARGEABLE lithium batteries can store more than twice as much energy per unit weight and volume as other rechargeable batteries^{1,2}. They contain lithium ions in an electrolyte, which shuttle back and forth between, and are intercalated by, the electrode materials. The first commercially successful rechargeable lithium battery³, introduced by the Sony Corporation in 1990, consists of a carbon-based negative electrode, layered LiCoO_2 as the positive electrode, and a non-aqueous liquid electrolyte. The high cost and toxicity of cobalt compounds, however, has prompted a search for alternative materials that intercalate lithium ions. One such is LiMn_2O_4 , which has been much studied as a positive electrode material⁴⁻⁷; the cost of manganese is less than 1% of that of cobalt, and it is less toxic. Here we report the synthesis and electrochemical performance of a new material, layered LiMnO_2 , which is structurally analogous to LiCoO_2 . The charge capacity of LiMnO_2 ($\sim 270 \text{ mA h g}^{-1}$) compares well with that of both LiCoO_2 and LiMn_2O_4 , and preliminary results indicate good stability over repeated charge-discharge cycles.

Many attempts have been made to prepare layered LiMnO_2 mainly involving the use of aqueous solutions⁸⁻¹⁰. The resulting products, though interesting, have stoichiometries which differ from LiMnO_2 , contain water or protons, are of poor crystallinity or do not maintain their structure during cycling. In contrast we have succeeded in preparing layered, anhydrous and stoichiometric LiMnO_2 which is analogous to LiCoO_2 and may be cycled in a rechargeable battery; it is obtained by ion exchange from NaMnO_2 . The sodium compound was synthesized by solid-state reaction between stoichiometric quantities of Na_2CO_3 and manganese (III) oxide at $700\text{--}730^\circ\text{C}$ for 18–72 hours under flowing argon¹¹. LiMnO_2 was obtained by refluxing NaMnO_2 with an excess of LiCl or LiBr in *n*-hexanol at $145\text{--}150^\circ\text{C}$ for 6–8 hours. After cooling to room temperature the product was filtered under suction and washed, first with *n*-hexanol and then with ethanol, and dried¹². Phase purity was established by powder X-ray diffraction.

The layered structure of LiMnO_2 was confirmed by powder neutron diffraction carried out on the POLARIS diffractometer at the ISIS pulsed source (Rutherford Appleton Laboratory) (Fig. 1a). The structure was refined by the Rietveld method using the program TF12LS based on the Cambridge Crystallographic Subroutine Library¹³. Adoption of a layered model yielded a final agreement factor, $R_{\text{weighted profile}}$, of 2.06% compared with an R_{expected} of 0.60%. A common structure obtained when attempts are made to prepare compounds with the LiMnO_2 composition is that of tetragonally distorted spinel $\text{Li}_2\text{Mn}_2\text{O}_4$ (ref. 4). This structure was tested in the refinement process; however, the fit obtained using the layered model was far superior to the fit assuming a tetragonal spinel structure for which $R_{\text{wp}} = 4.79\%$. The layered structure of LiMnO_2 is shown in Fig. 1b. The oxide ions are arranged in close-packed layers which are stacked in an ABC repeat sequence, that is, cubic close packing. Manganese ions are located in each octahedral site between the first and second oxide layers (Fig. 1b). Between the second and third layers the Li^+ ions reside also in octahedral sites. Refinement was carried out permitting the Li^+ and Mn^{3+} ions to occupy their

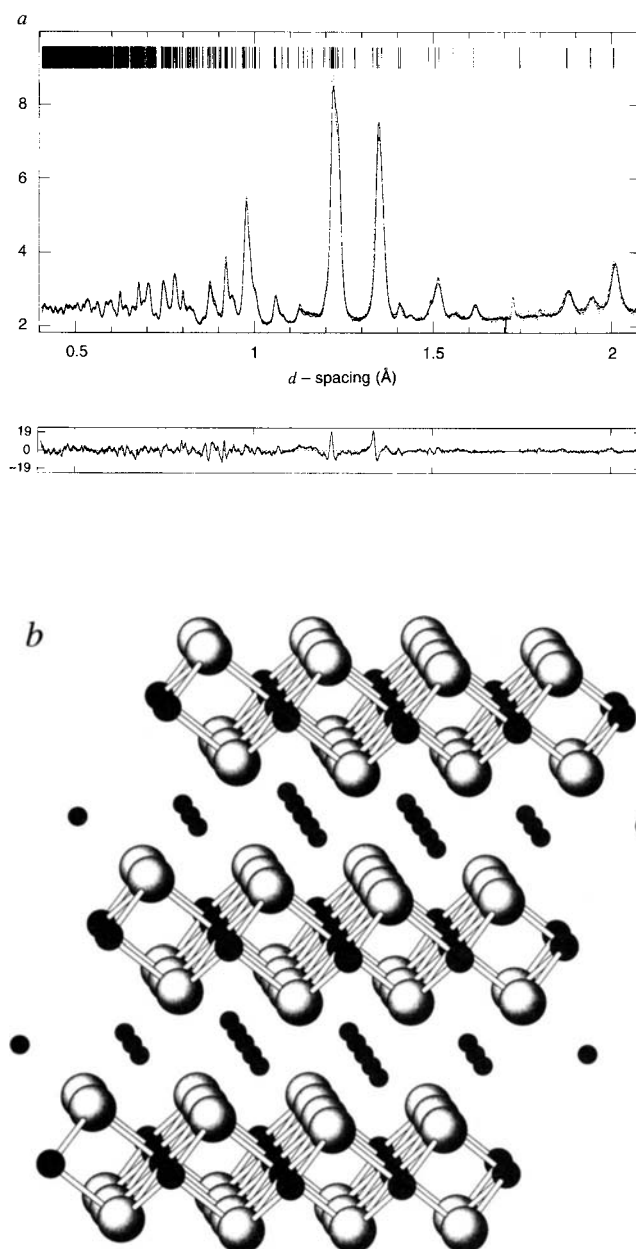


FIG. 1 a, Observed (dots) and calculated (solid line) neutron-diffraction profiles for monoclinic LiMnO_2 . The lower plot shows the difference/estimated standard deviation b, Structure of LiMnO_2 emphasising the layered nature of the material. Mn, large dark circles; Li, smaller dark circles; O, pale circles.

TABLE 1 Crystallographic parameters of LiMnO_2

Atom	Wyckoff symbol	x	y	z	B_{eq}	Site occupancy
Li1/Mn1	2d	0	0.5	0.5	2.4(2)	0.91/0.09(4)
Li2/Mn2	2a	0	0	0	0.72(6)	0.10/0.90(3)
O1	4i	0.2723(3)	0	0.7706(2)	0.68(4)	1

Monoclinic, space group $C2/m$ (no. 12). Unit-cell dimensions: $a = 5.4387(7) \text{ \AA}$, $b = 2.80857(4) \text{ \AA}$, $c = 5.3878(6) \text{ \AA}$, $\beta = 116.006(3)^\circ$, $\chi^2 = 11.83$. $R_{\text{exp}} = 0.60\%$, $R_p = 1.86\%$, $R_{\text{wp}} = 2.06\%$, $R_1 = 3.98\%$.

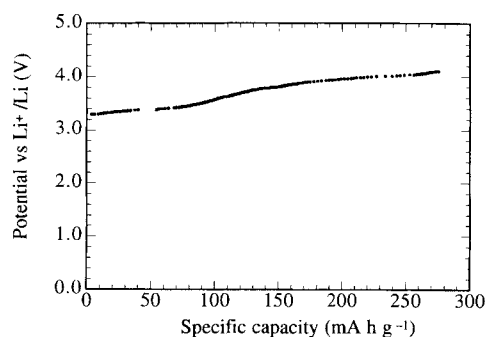


FIG. 2 Variation of electrode potential with capacity on charging LiMnO_2 at a current density of $10 \mu\text{A cm}^{-2}$.

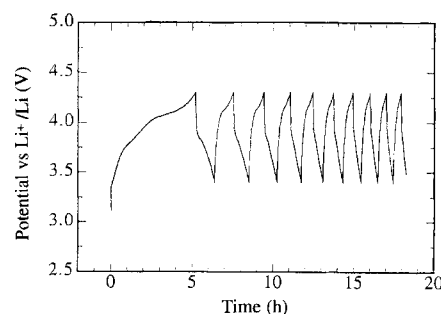


FIG. 3 Cycling of the LiMnO_2 electrode at a current density of 0.5 mA cm^{-2} and between voltage limits of 3.4 to 4.3 V vs Li^+/Li .

own sites and those of each other. The Li^+ and Mn^{3+} ion occupancies were free to refine on both sites, the only constraint being that the total occupancy of each site was set to equal one. The total lithium and manganese contents were respectively 1.02(7) and 0.98(7), consistent with the LiMnO_2 stoichiometry. Atomic absorption analysis was carried out (Unicam PU9400X) and provided further confirmation of the stoichiometry. The maximum occupancies of Li on the Mn sites (2a) and Mn on the Li (2d) sites were respectively 0.10(3) and 0.09(4), demonstrating that the lithium and manganese ions reside predominantly on their own sites. Due to the strong Jahn–Teller nature of high-spin Mn^{3+} ($3d^4$), the local site symmetry around the Mn^{3+} ions is distorted somewhat from a regular octahedron and the crystal structure is not rhombohedral but monoclinic (space group $C2/m$). The crystallographic data are given in Table 1.

Interest in layered LiMnO_2 as a positive electrode in rechargeable lithium batteries stems from the fact that Li^+ ions and electrons may be removed and reinserted into this compound, that is, it is an intercalation host for lithium. The electrochemical performance of LiMnO_2 was investigated with a three-electrode cell composed of lithium metal counter and reference electrodes, the working electrode being fabricated by compressing powdered LiMnO_2 (80%), carbon black (13.3%) and PTFE (6.7%) on to a metal grid. The electrolyte consisted of a 1 M solution of LiClO_4 dissolved in propylene carbonate. The salt was rigorously dried by heating under vacuum at 150°C and the solvent was distilled as described elsewhere¹⁴. The cell was subjected to charging at a current density of $10 \mu\text{A cm}^{-2}$. The resulting voltage curve is shown in Fig. 2. These data confirm that up to 0.95 lithiums per formula unit may be extracted from LiMnO_2 on initial charging, corresponding to a capacity of 270 mA h g^{-1} . Preliminary measurements at the much higher current density of 0.5 mA cm^{-2} suggest that a capacity approaching 200 mA h g^{-1} may be obtained up to a cut-off potential of 4.3 V.

The earliest lithium batteries, introduced in the 1980s, used lithium metal as the negative electrode. Despite some problems with flammability, there is still interest in using such cells for electric-vehicle applications because of their high gravimetric energy density. For such cells cathodes with a potential of 3 V are sufficient. When combined with carbon-based anodes in 'rocking-chair' cells (in which both electrodes form intercalation

compounds), cathodes with potentials of more than 3 V are required; however the move from petroleum coke to graphite has resulted in an anode with a potential closer to that of lithium metal over a greater range of composition. A practical capacity of $\sim 150 \text{ mA h g}^{-1}$ is obtainable from the cathode used in the present generation of commercially available rechargeable lithium batteries, that is, LiCoO_2 . A slightly smaller practical capacity may be obtained at 3 and 4 volts from LiMn_2O_4 . The capacities of 200 and 270 mA h g^{-1} obtained on initially charging LiMnO_2 and already mentioned above, demonstrate that this compound is both scientifically interesting and potentially attractive as a cathode for rechargeable lithium batteries. However a full analysis of the utility of this cathode in comparison with others will require extensive measurements of the capacity variation on cycling at different current densities and over different voltage ranges. Some preliminary cycling data are presented below.

X-ray diffraction data reveal that for $(1-x) \approx 0.5$ in $\text{Li}_{1-x}\text{MnO}_2$ the structure is rhombohedral. On removal of lithium from LiMnO_2 , the Jahn–Teller active Mn^{3+} ion is oxidized to Mn^{4+} with the resulting loss of the monoclinic distortion. The layered compound also exhibits a voltage transition at a composition of $\sim \text{Li}_{0.5}\text{MnO}_2$. This is the same composition (LiMn_2O_4) at which lithium manganese oxide with the spinel structure transforms from a 3-V to 4-V cathode, suggesting that the voltage transition is not related specifically to the spinel structure. However, a full discussion of these more detailed aspects must await the availability of further structural data.

Some preliminary cycling data are presented in Fig. 3. The cell was cycled at a constant current density of 0.5 mA cm^{-2} between the potential limits 3.4 and 4.3 V. Previous studies using cyclic voltammetry indicated that the electrolyte was stable in contact with this electrode up to at least 4.4 V. X-ray diffraction carried out on the layered material at different stages of cycling indicate that the layered structure, is retained during lithium removal and reinsertion. Although the capacity declines on cycling, it must be stressed that the voltage range has not been optimized and includes both voltage plateaux. In the case of the LiMn_2O_4 spinel, cycling over both plateaux results in a greater capacity fade than if confined to just one plateau. It should also be recalled that the early spinel materials showed very poor cyclability, and only later optimization yielded satisfactory results. □

Received 14 February; accepted 2 May 1996.

- Oyama, N., Tatsuma, T., Sato, T. & Sotomura, T. *Nature* **373**, 598–600 (1995).
- Scrosati, B. *Nature* **373**, 557–558 (1995).
- Nagaura, T. 3rd Int. Battery Seminar (Deerfield Beach, FL, 1990).
- Thackeray, M. M., David, W. I. F., Bruce, P. G. & Goodenough, J. B. *Mater. Res. Bull.* **18**, 461–472 (1983).
- Huang, H. & Bruce, P. G. *J. Power Sources* **54**, 52–57 (1995).
- Tarascon, J.-M. & Guyomard, D. *Electrochimica Acta* **38**, 1221–1231 (1993).
- Pistoia, G. & Wang, G. *Solid St. Ionics* **66**, 135–142 (1993).

- Rossouw, M. H., Liles, D. C. & Thackeray, M. M. *J. Solid St. Chem.* **104**, 464–466 (1993).
- Thackeray, M. M. *J. electrochem. Soc.* **142**, 2558–2563 (1995).
- Leroux, F., Guyomard, D. & Piffard, Y. *Solid St. Ionics* **80**, 299–306 (1995).
- Fuchs, B. & Kemmler-Sack, S. *Solid St. Ionics* **68**, 279–295 (1994).
- Bruce, P. G. & Armstrong, A. R. UK Patent Application January 1996.
- Matthewman, J. C., Thompson, P. & Brown, P. J. *J. appl. Crystallogr.* **15**, 167–173 (1982).
- Huang, H. & Bruce, P. G. *J. electrochem. Soc.* **141**, L106–L107 (1994).

ACKNOWLEDGEMENTS. We thank H. Huang for her assistance with the electrochemical measurements. P.G.B. is grateful to the EPSRC for financial support.

Characterization of a cage form of the water hexamer

K. Liu*, M. G. Brown*, C. Carter*, R. J. Saykally*, J. K. Gregory† & D. C. Clary‡

* Department of Chemistry, University of California, Berkeley, California 94720-1460, USA

† Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK

‡ Department of Chemistry, University College London, London WC1H 0AJ, UK

WATER has been studied more extensively than any other liquid, yet its microscopic properties remain poorly understood. The difficulty in obtaining a rigorous molecular-scale description of water structure is largely a consequence of the extended, dynamic hydrogen-bonded network that exists throughout the liquid¹. Studies of the structure and dynamics of isolated small clusters of water molecules^{2–6} provide a means of quantifying the intermolecular forces and hydrogen-bond rearrangements that occur in condensed phases. Experiments^{2–7} and theory⁸ strongly suggest that the water trimer, tetramer and pentamer have cyclic minimum energy structures. Larger water clusters are expected⁸ to have three-dimensional geometries, with the hexamer representing the transition from cyclic to such three-dimensional structures. Here we report investigations by terahertz laser vibration–rotation tunnelling spectroscopy³ of the structure of the water hexamer. A comparison of our results with quantum Monte Carlo simulations of this species suggests that the most stable form of $(\text{H}_2\text{O})_6$ is indeed a cage-like structure, held together by eight hydrogen bonds (Fig. 1).

We have measured a perpendicular vibration–rotation–tunnelling (VRT) hybrid (both b- and c-type) band of an $(\text{H}_2\text{O})_n$ cluster (Fig. 2) centred at 2.491199 THz (83.09744 cm^{-1}) using the Berkeley terahertz laser spectrometers⁹. The size of the water cluster was determined from an isotope mixture test described previously⁴, wherein the absorption line intensity was carefully measured as a function of the H mole fraction in the $\text{H}_2\text{O} + \text{D}_2\text{O}$ mixture employed for the production of water clusters in a pulsed supersonic slit jet¹⁰. The size of the cluster thus determined was 5.99 ± 0.15 to 95% confidence, establishing the spectral carrier as the hexamer.

The highest level *ab initio* calculations yet performed for the hexamer reveal the existence of several low-lying structures¹¹, but predict that the minimum-energy form is the three-dimensional cage shown in Fig. 1 (ref. 8). The experimental rotational constants extracted from the present VRT spectra agree most closely with those predicted for this cage structure (M. Pedulla and K. D. Jordan, personal communication). These *ab initio* studies also predict that zero-point vibrational effects can alter the energy ordering of the low-lying hexamer structures. As described below, a rigorous first-principles quantum-mechanical investigation of the relative stability of these hexamer conformers employing a realistic potential model for water validates this prediction and likewise suggests the cage structure as the most stable form of $(\text{H}_2\text{O})_6$. Furthermore, this model potential indicates that four non-equivalent but nearly degenerate cage structures are linked by successive flips of the two free hydrogens which reside on the doubly bonded monomers acting as both donor and acceptor (Fig. 1) (D. J. Wales, personal communication). This low-barrier flipping motion is likely to be the origin of the vibration observed in this work.

We have also observed quantum tunnelling splittings (Fig. 2) for each vibration–rotation transition. The distinctive triplet (intensity ratios = 9 : 6 : 1) spectral pattern is consistent with a hydrogen-bond rearrangement that involves the exchange (an effective monomer C_2 rotation) of the bonded and free hydrogens

within two (and only two) slightly inequivalent monomer units; in the cage structure (Fig. 1), the two doubly bonded monomers are likely to have sufficiently low hydrogen-exchange barriers to yield such resolvable tunnelling splittings (see Fig. 1 legend).

The potential energy of the water hexamer used here is defined by means of a rigid monomer intermolecular potential-energy surface, based on the model first described by Millot and Stone¹², and includes many-body forces¹³. This model potential, although not definitive, predicts low-energy structures for water clusters in good agreement with the best electronic structure (*ab initio*) calculations and rotational constants for the dimer, trimer, tetramer and pentamer that agree well with experiment. A large number of energy minimizations starting from various initial geometries yielded five low-energy water hexamer structures. In identifying these low-energy structures, we considered only the lowest-energy forms of any geometrically similar species and were also guided by previous work^{8,11}. The diffusion quantum Monte

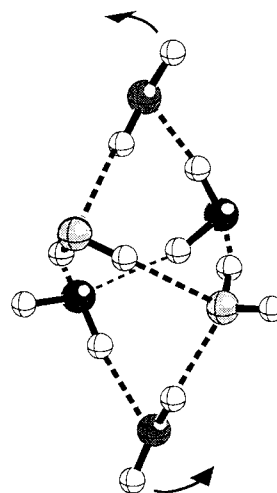


FIG. 1 The cage-like $(\text{H}_2\text{O})_6$ equilibrium structure in best agreement with the measured rotational constants. The facile large-amplitude 'flipping' motions (shown by curved arrows) of the free O–H bonds on the two single-donor single-acceptor monomers are analogous to the flipping motions found in the cyclic trimer⁴, tetramer⁵ and pentamer⁶. These induce a dipole moment change that is perpendicular to the long symmetry axis of the molecular inertial ellipsoid (a-axis for a near-prolate rotor), and are likely to be the origin of the perpendicular type (both b- and c-type) transitions observed here. (These band types correspond to dipole moment changes along the b and c principal inertial axes of a near prolate rotor, wherein the a-axis is the approximate symmetry axis.) Note that hydrogen-exchange tunnelling motions within the same two monomers could also produce the observed distinctive triplet spectral pattern, as explained below. The cluster VRT states correlate to the monomer nuclear spin wavefunctions of ortho (o) and para (p) symmetries. A stepwise tunnelling performed by each monomer with the corresponding splittings β_1 and β_2 leads to the VRT states which are labelled by the o–o (9), o–p (3), p–o (3), and p–p (1) symmetries with the uneven respective spacings β_2 , $\beta_1 - \beta_2$ and β_2 , and the statistical weights given in parentheses. Two factors contribute to the observed pattern being a triplet instead of a quartet. First, the very similar tunnelling barrier heights for the two single-donor/single-acceptor monomers conceivably produce only slightly different β_1 and β_2 values. Second, the measured line spacing represent only the change in tunnelling splittings between the ground (double-primed) and excited (single-primed) states, that is, $\beta'_2 - \beta''_2$, $(\beta'_1 - \beta'_2) - (\beta''_1 - \beta''_2)$, and $\beta'_2 - \beta'_2$. The small value of $(\beta'_1 - \beta'_2) - (\beta''_1 - \beta''_2)$ is probably responsible for the o–p and p–o states collapsing into an unresolvable singlet with a combined spin weight of 6 (= 3 + 3), thereby yielding a triplet pattern with intensity ratios of 9:6:1. The boat structure (see Fig. 2) would also yield a triplet with intensity ratios of 9:6:1, as the two single-donor/single-acceptor monomers become equivalent and the tunnelling splittings from the effective C_2 rotation of each monomer therefore become identical. But the energetic considerations discussed in the text and more importantly the experimentally determined structural parameters (Table 1) argue in favour of the cage, rather than the boat, assignment.

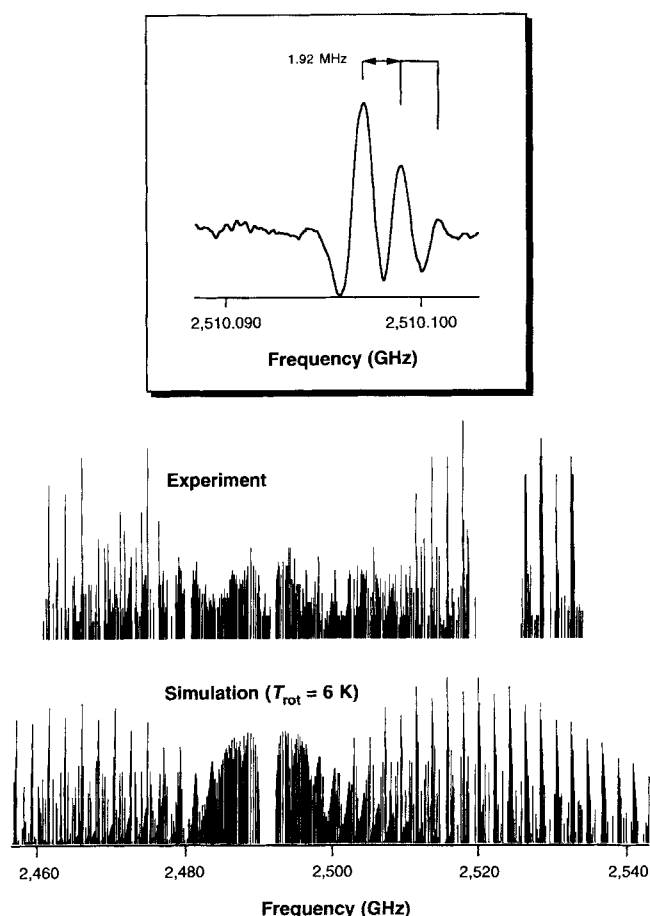


FIG. 2 The observed and simulated VRT spectra of the near-prolate $(\text{H}_2\text{O})_6$ cluster. Only the regions actually scanned in the experiment are shown here. Inset, the actual experimental spectrum of the resolved tunnelling triplet (equal spacings of 1.92 MHz) splitting pattern that is associated with (and invariant with) each rovibrational transition. The intensity simulation incorporated the transition moment and a Boltzmann distribution at a rotational temperature of 6K. The line positions were calculated with the molecular constants obtained from a nonlinear least-squares fit of the observed transition frequencies to an S-reduced Watson hamiltonian, which determined individual A , B and C rotational constants for the observed perpendicular band. The measured rotational constants, determined through a least-squares fit of 265 VRT transitions, are (in MHz) $A'' = 2,163.61(80)$, $B'' = 1,131.20(62)$, $C'' = 1,068.80(62)$ for the vibrational ground state and $A' = 2,156.10(80)$, $B' = 1,129.77(61)$, $C' = 1,065.53(61)$ for the vibrationally excited state (the 1σ uncertainty in the last digits of the parameters are given in the parentheses). Jet-cooled water clusters were produced and cooled to 6K in a pulsed planar supersonic expansion¹⁰ of Ar bubbled through room-temperature water, and studied by means of direct absorption with the Berkeley tunable terahertz laser spectrometer described elsewhere⁹.

TABLE 1 Calculated and experimental ground-state rotational constants for the water hexamer (in MHz)

	At potential minimum			DQMC* (vibrationally averaged)		
	A	B	C	A	B	C
Cyclic	1,269	1,250	636	1,211	1,211	598
Book	1,984	1,173	866	1,798	1,078	802
Boat	2,015	1,067	999	1,733	1,125	1,054
Prism	1,752	1,504	1,421	1,607	1,355	1,256
Cage	2,345	1,195	1,145	2,136	1,096	1,043
Expt	—	—	—	2,163.61(80)	1,131.20(62)	1,068.80(62)

* Calculated by the diffusion quantum Monte Carlo method.

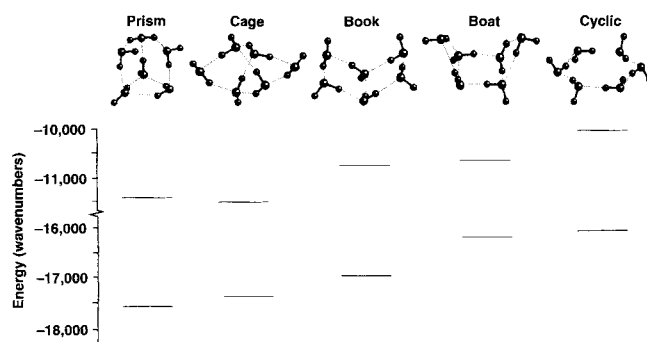


FIG. 3 Theoretical predictions of the stabilities of the five lowest-energy water hexamer structures obtained in the DQMC calculations with the model potential. Values of D_e (lower line) and D_0 (upper line) are shown. The zero-point energy is equal to $D_0 - D_e$. Note that the cage structure has a D_e smaller than that for the prism structure by 213 cm^{-1} , whereas the D_0 calculations suggest that the cage structure is more stable by 62 cm^{-1} (89K).

Carlo (DQMC) method^{14–16} was employed to calculate rigorously the effect of the large-amplitude intermolecular motions on the structures and properties of the hexamer. This technique solves the nuclear Schrödinger equation accurately by exploiting its isomorphism with a modified diffusion equation, and in this case it involves 30 strongly coupled degrees of freedom.

Results of the calculations using energy minimizations and DQMC for the five structures are shown in Fig. 3. A striking finding is that although a prism structure has the lowest equilibrium dissociation energy (D_e), the cage structure becomes slightly more stable when quantum vibrational zero-point energy corrections are properly taken into account (D_0). We have calculated rotational constants for the five structures, both for the minimum-energy and vibrationally averaged structures. The effect of the vibrational averaging is dramatic, with the A , B and C rotational constants (inverse moments of inertia) changing by 10–20% (see Table 1). The cage structure has vibrationally averaged rotational constants corresponding to a near-prolate top which, as can be seen from Table 1, are in close agreement with experiment. None of the other four low-energy structures produced a reasonable set of rotational constants. Hence we conclude that the VRT spectrum described here is indeed that of this cage form of the water hexamer, and that the large intermolecular zero-point energies of the hydrogen bonds are crucial in producing this as the minimum-energy structure. The fact that both high-level *ab initio* calculations and the DQMC results predict the cage as the lowest-energy form, coupled with observations that no structures other than the most stable have ever been detected by VRT spectroscopy for any cluster in 6K argon supersonic jets³, is fairly compelling (but not yet definitive) evidence that the cage is indeed the most stable form of water hexamer.

The water trimer, tetramer, pentamer and hexamer are some of the dominant structures identified in room-temperature liquid water^{17,18}. Our DQMC calculations give an average O–O bond distance of 2.85 Å for the cage hexamer, which is very close to the value of 2.85 Å reported for liquid water at 298 K (ref. 19). Furthermore, the DMC O–O bond distance is 2.76 Å for the cyclic isomer of the hexamer, which is very near the value of 2.759 Å for ice (I_h) at 223 K (refs 20, 21). These comparisons illustrate clearly the special role that the isomers of the hexamer have in relation to the properties of liquid water and ice. From these and future investigations of water clusters, we can ultimately hope to extract a quantitative description of the many-body interactions which are responsible for the complexity of liquid water's behaviour. □

Received 19 February, accepted 22 April 1996.

1. Luzar, A. & Chandler, D. *Nature* **379**, 55–57 (1996).
2. Liu, K., Cruzan, J. D. & Saykally, R. J. *Science* **271**, 929–933 (1996).
3. Saykally, R. J. & Blake, G. A. *Science* **259**, 1570–1575 (1993).
4. Pugliano, N. & Saykally, R. J. *Science* **257**, 1937–1940 (1992).
5. Cruzan, J. D. et al. *Science* **271**, 59–62 (1996).
6. Liu, K., Brown, M. G., Cruzan, J. D. & Saykally, R. J. *Science* **271**, 62–64 (1996).
7. Pribble, R. N. & Zwier, Z. S. *Science* **265**, 75–79 (1994).
8. Kim, K., Jordan, K. D. & Zwier, T. S. *J. Am. chem. Soc.* **116**, 11568–11569 (1994).
9. Blake, G. A. et al. *Rev. Sci. Instrum.* **62**, 1701–1716 (1991).
10. Liu, K. et al. *Rev. Sci. Instrum.* **67**, 410–416 (1996).
11. Tsai, C. J. & Jordan, K. D. *Chem. Phys. Lett.* **213**, 181–188 (1993).
12. Millot, C. & Stone, A. J. *Molec. Phys.* **77**, 439–462 (1992).

13. Gregory, J. K. & Clary, D. C. *J. chem. Phys.* **103**, 8924–8930 (1995).
14. Anderson, J. B. *J. chem. Phys.* **63**, 1499–1503 (1986).
15. Coker, D. F. & Watts, R. O. *Molec. Phys.* **58**, 1113–1123 (1986).
16. Buch, V. J. *J. chem. Phys.* **97**, 726–729 (1992).
17. Speedy, R. J., Madura, J. D. & Jorgensen, W. L. *J. phys. Chem.* **91**, 909–913 (1987).
18. Belch, A. C. & Rice, S. A. *J. Chem. Phys.* **86**, 5676–5682 (1987).
19. Narten, A. H., Thiessen, W. E. & Blum, L. *Science* **217**, 1033–1034 (1982).
20. Peterson, S. W. & Levy, H. A. *Acta Crystallogr.* **10**, 70–76 (1957).
21. Kuhs, W. F. & Lehmann, M. S. *J. phys. Chem.* **87**, 4312–4313 (1983).

ACKNOWLEDGEMENTS. We thank J. G. Loeser and D. J. Wales for helpful discussions. The experimental work was supported by the Experimental Physical Chemistry Program of the US National Science Foundation; the theoretical work was funded by the UK Engineering and Physical Sciences Research Council.

Role of orbitally induced changes in tundra area in the onset of glaciation

R. G. Gallimore & J. E. Kutzbach

Center for Climatic Research, Institute for Environmental Studies,
University of Wisconsin-Madison, 1225 West Dayton Street, Madison,
Wisconsin 53706, USA

THE link between glacial–interglacial cycles and changes in insolation due to variations in the Earth's orbital parameters is well established^{1–4}. But of the attempts to simulate incipient glaciation using three-dimensional general circulation models (GCMs) driven by orbital forcing alone^{5–10}, only one¹⁰ has been successful. GCM experiments^{7,11} show that reduced summer insolation 115,000 years ago (during an interglacial-to-glacial climate shift) produces sufficient high-latitude cooling to cause expansion of tundra at the expense of boreal forest¹¹, which in turn can induce more cooling^{11–14}. Here we show, using a global climate model, that the increase in surface albedo (under snow-covered conditions) that results from a biome model estimate¹¹ of tundra expansion 115,000 years ago is sufficient to induce glaciation over extreme-northeastern Canada. If the additional cooling from this estimated tundra expansion induces further expansion, then widespread glaciation occurs at latitudes above 65° N. These results suggest that the climate feedback from high-latitude tundra expansion might have contributed to the onset of the most recent glaciation.

We performed a set of five 25-year seasonal cycle experiments with the NCAR atmospheric GCM (Community Climate Model version 1; CCM1) coupled to a simple mixed-layer ocean with interactive sea ice^{11,15}: Control (modern climate simulation), Exp-A (climate simulation with orbital parameters for around 115,000 yr BP, modern CO₂ and surface albedo), Exp-B (climate simulation with orbital parameters and CO₂ concentration for 115,000 yr BP but modern surface albedo), Exp-C (climate simulation with orbital parameters and CO₂ concentration for 115,000 yr BP and a modest increase in surface albedo representing the effects of orbitally forced expanded tundra under snow-covered conditions) and Exp-D (the same as Exp-C except for a larger increase in surface albedo representing the feedback of further expansion of tundra at the expense of the boreal forest induced by the cooling in Exp-C). Results are presented for the final 12 years of quasi-equilibrium conditions. The influence of the individual factors on initiation of high-latitude glaciation is examined by taking the differences between experiments: orbital parameter change (Exp-A minus Control), CO₂ change (Exp-B minus Exp-A), albedo increase with modest tundra expansion (Exp-C minus Exp-B), and albedo increase with larger tundra expansion (Exp-D minus Exp-C). The combined influences are examined by taking differences between the appropriate experiment and the control.

In Exp-A the orbital parameters are changed from modern values to an idealized set of conditions approximating those around 115,000 yr BP (eccentricity from 0.0167 to 0.04, axial tilt from 23.45° to 22.45° and perihelion from 3 January to 21 December). The combination of decreased axial tilt and increased eccentricity produces an 8% decrease in Northern Hemisphere summer insolation compared with modern values. In high northern latitudes the decreased summer insolation promotes colder conditions compared with the present^{7,8,11}.

Both Control (modern) and Exp-A used an atmospheric CO₂ concentration of 330 p.p.m. In Exp-B the CO₂ concentration is reduced from 330 to 267 p.p.m. based on evidence from ice cores¹⁶. Numerous GCM experiments indicate that a decrease in atmospheric CO₂ initiates surface and tropospheric cooling by the reduced emission and absorption of long-wave radiation by CO₂ and subsequent feedbacks involving water vapour, clouds, snow and sea ice^{17,18}.

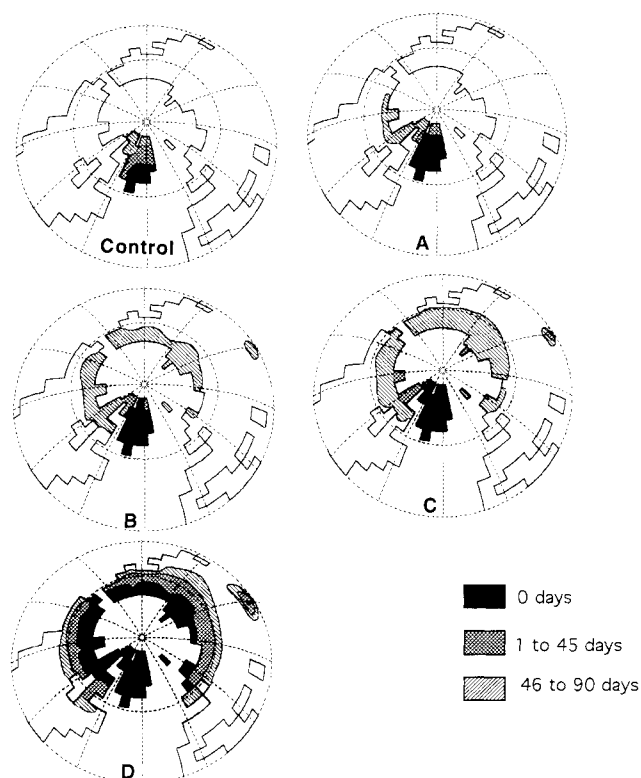


FIG. 1 Northern Hemisphere distribution of simulated 12-year mean snowcover-free duration on model grid for modern Control and Exp-A to Exp-D. Three areas are delineated: snow-free duration between 46 and 90 days (hatched), snow-free duration between 1 and 45 days (cross-hatched), and permanent snow cover (solid). Within the area of 1–45 days snow-free duration, there are locations that experienced occasional years with year-round snowcover.

Observational evidence indicates widespread changes in vegetation at the end of the previous interglacial (115,000 yr BP) that are consistent with an expansion of tundra/steppe at the expense of forest at high latitudes^{19–21}. A shift from forest to steppe occurred over Europe²⁰ and a shift to cooler vegetation types is found over eastern Canada²¹. The total area of observed high-latitude tundra/steppe expansion is not known, in part because evidence in some locations has probably been destroyed by landscape changes associated with the subsequent glaciation. We have inferred the potential size of tundra expansion from model studies¹¹. An application of a biome model to the results of Exp-A suggests a 25% areal expansion of tundra (from modern) occurred poleward of 60° N as a consequence of the summer cooling induced by the 115,000-yr BP orbital conditions¹¹. The southward expansion, averaging about 5° of latitude, occurs over most longitudes but is most marked in Siberia and North America. The effect of the expanded tundra (relative to forest) is to enhance the local surface albedo by ~0.3–0.4 under snow-covered conditions^{12,22} and to decrease surface temperatures significantly in spring and early summer¹³. The lower albedo for forested land results from the protrusion of a low-reflectance tree canopy above a highly reflective snow-covered ground. As a consequence, the fraction of area with a highly reflective snow cover (snowcover fraction) is considerably lower for land covered with boreal forest than tundra vegetation.

The CCM1 calculates surface albedo by using a single snowcover fraction. Because the model snowcover fraction is relatively high (0.8) in the control, we could not sufficiently increase the surface albedo simply by increasing the snowcover fraction for those grid points that switched from boreal forest to tundra. Moreover, given the coarseness of model resolution (4.5° latitude by 7.5° longitude) and the large zonal extent of southward tundra expansion simulated by the biome model, the sense of the snowcover response to tundra expansion is approximated by distributing a relatively small increase in the albedo of snow-covered land (0.08) over a large area rather than a large albedo increase (0.30) over a relatively small area (the 25% area of expanded tundra from the biome simulation). For Exp-C, this approximation is accomplished by increasing the snowcover fraction from 0.8 to 0.9 at all land points poleward of 60° N (excluding Greenland). This increases the surface albedo from 0.63 to 0.71 and has the effect of reducing the radiation available to melt snow poleward of 60° N by nearly the same amount as would occur with 25% of the area having a 0.3 albedo increase and 75% of the area having no increase.

In Exp-D the surface albedo poleward of 60° N was raised an additional 0.08 by a further increase in snowcover fraction. This additional increase in surface albedo was designed to represent a plausible upper bound accounting for a feedback effect of an additional 25% expansion of tundra produced by the cooling in Exp-C. Iterative coupling of a biome model with an atmospheric climate model supports this argument for additional tundra expansion from the initial estimate used in Exp-C (N. de Noblet, personal communication).

The individual effects of the orbital, CO₂ and modest expanded tundra forcing produced comparable climate changes (from

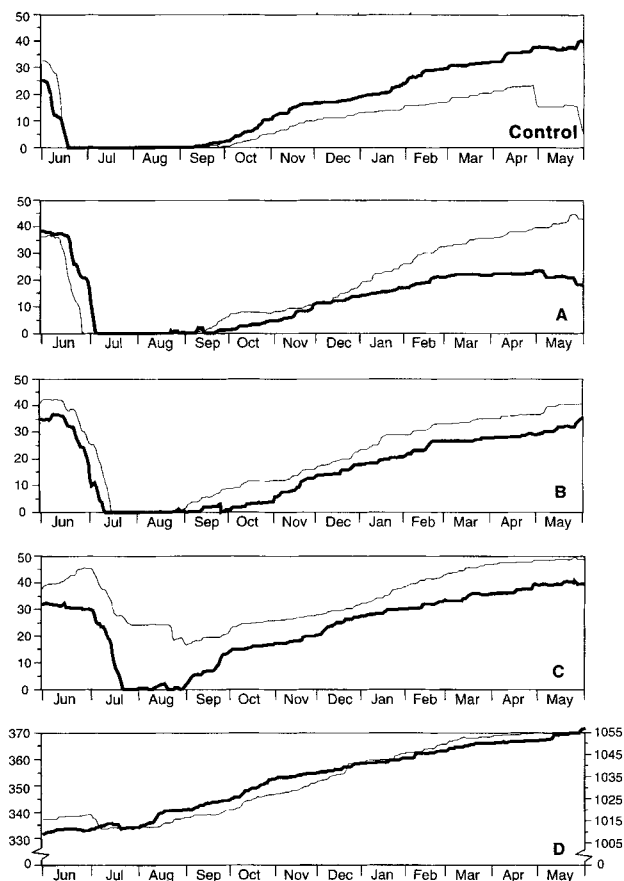


FIG. 2 Simulated daily snow depth (cm water equivalent) from June to May for the year with shortest snow-free season over model grid cells centred on northeast Keewatin (69°N, 90°W) (heavy curve) and north Baffin Island (73°N, 83°W) (light curve) in modern Control and Exp-A to Exp-C. Snow cover is permanent for both these regions in Exp-D and is given for model year 12 (left scale for Keewatin and right scale for north Baffin Island).

modern) over the high-latitude northern continents in summer (Table 1). Each forcing effect induced surface temperatures to decrease by ~2.5–3.0 °C. The summer cooling in each case leads to progressively greater summer snowfalls and shortening of the snowcover-free season by 1–3 weeks. With each added forcing effect, additional expansion and thickening of sea ice occurred over the polar ocean. Together, the orbital and CO₂ forcing induce ocean surface temperature to decrease by 2–4 °C in summer over Hudson Bay and the Norwegian and Labrador seas. This temperature response, although significant, is less than half that simulated in a parallel experiment with a different model¹⁰.

The combination of orbital, CO₂, and modest expanded tundra

TABLE 1 Twelve-year summer mean (June–August) of simulated climate variables over land averaged from 60° to 90° N.

	Control	Exp-A	Exp-B	Exp-C	Exp-D
North America					
Surface temperature (°C)	12.6	9.9	7.5	4.6	–10.5
Snowfall rate (mm H ₂ O d ^{–1})	0.00	0.05	0.11	0.20	1.03
Days of year with snowcover	237	249	269	287	354
Europe–Asia					
Surface temperature (°C)	14.3	11.2	9.1	6.3	–8.0
Snowfall rate (mm H ₂ O d ^{–1})	0.00	0.03	0.07	0.16	0.87
Days of year with snowcover	240	247	264	282	348

forcings (Exp-C) acts to decrease the snowcover-free season by 6–7 weeks over northern Eurasia and North America. The average summer temperature, although above freezing, is 8 °C lower than modern. The greater tundra expansion in Exp-D than in Exp-C leads to a catastrophic climate change over the high-latitude northern continents. Summer temperatures fall an additional 15 °C to well below freezing. Much of the region poleward of 60° N has permanent snow cover, with the snowcover-free season over northern North America and Eurasia further reduced by 9 weeks to about 1–2 weeks. A substantial fraction of summer precipitation now occurs as snow.

The effect of each sequential change of forcing on snow cover is readily seen by following the areal expansion of a region with a specified interval of snowcover-free duration (Fig. 1). The region of 0–90-day snow-free duration, confined in the control to the area around Greenland, expands over northern Canada with the orbital forcing change and then to northern Eurasia with the combined orbital and CO₂ forcing changes. The additive influence of the modest increased tundra forcing leads to small areas of northern Canada, Tibet and extreme northern Eurasia with snowcover-free seasons between 1 and 45 days. Permanent snow cover, confined to the Greenland ice cap in Control and Exp-A and Exp-B, now extends to north Ellesmere Island. A few places in the areas encompassed by 1–45-day mean snow-free duration (north Baffin Island, south Ellesmere Island, Devon Island and Victoria Island) are close to incipient glaciation because they have year-round snowfields for some model years (1–2 and 4–5). Observational evidence points to north Baffin Island and northeast Keewatin as likely sites for initiation of the Laurentide ice sheet at the end of the last interglacial²³.

The additional expansion of tundra in Exp-D produces an extensive area of permanent snow cover over the northern continents poleward of 65° N and a small area over Tibet (Fig. 1). The area of permanent snow cover encompasses regions of major early glaciation (Baffin Island, Keewatin, Fenno-Scandinavia) as well as areas either not extensively glaciated (for example, Alaska) or where uncertainty exists about early glaciation (for example, Siberia, Tibet)^{10,23–25}. Snow depths increase steadily by ~30–50 cm of water equivalent per year, which is sufficient to generate a glacial ice volume after 6,000–8,000 years that would lower global-average sea level by over 70 m. Palaeoclimate data indicate that in the roughly 8,000 years from ~120,000 to 112,000 yr BP the increased global ice volume and associated sea level lowering was about one-third of that between today and the Last Glacial Maximum (LGM) at 21,000 yr BP^{26,27}. Based on the most recent estimate of sea level lowering for the LGM (105 m)²⁵, the observed reduction in sea level between 120,000 and 112,000 yr BP is ~40 m, the implied sea level lowering of 70 m in Exp-D seems excessive. But our calculation of sea level lowering does not consider possible orographic modification of snow accumulation or ice-sheet dynamical processes that would act to limit the ice buildup implied by the model's snow accumulation rate (for example, bedrock dynamics, ice flow and calving into icebergs)^{14,26}.

Two mechanisms contribute to the growth of summer snow cover within the experiments (Fig. 2). First, with the addition of each forcing component, the previous winter snow pack extends progressively longer into the summer season as a result of the higher surface albedo and thus the decreased spring–early-summer snow melt. The second mechanism involves progressively earlier occurrences of late summer snowfalls. The combination of cool summer orbital forcing and lowered CO₂ levels for 115,000 yr BP and the modest expansion of tundra (Exp-C) reduces temperatures sufficiently to induce snowfalls throughout the summer season over northeast Canada. These two mechanisms act together to produce permanent summer snowcover over Ellesmere Island (not shown), occasional year-round snowfields over north Baffin Island (Fig. 2) and some years with only a short snow-free season (three weeks) over northeast Keewatin (Fig. 2). A further tundra expansion (Exp-D) produces permanent snow cover over a broad area of high-latitude land (Fig. 1, 2).

The results of our experiments strongly suggest that a vegetative feedback of expanded tundra was important to the initiation of the last glacial age. This conclusion is supported by the fact that only one GCM study¹⁰ out of a number of attempts^{5,7–10} has been able to induce incipient glaciation with just orbital and CO₂ forcing for ~115,000 yr BP. Further insight into reasons for the success or failure of one model over another to produce incipient glaciation requires understanding of the effects of model biases. Differences in model resolution could certainly be a factor¹⁰. Another factor could be differences between models in their treatments of ocean, sea ice, and land surface processes. Still another factor could be differences in parametrization of snow hydrology, leading to different biases in snow climatology in the control climate. CCM1 is known to have biases in the simulation of high-latitude climate that include excessive precipitation and winter snow accumulation²⁸. In spite of this bias the model simulates a snowcover-free season over northeast Keewatin (70° N) of ~15 weeks, or ~6 weeks longer than observed (8–10 weeks), because of an overactive snowmelt mechanism⁷. If the CCM1 were not biased toward this longer than observed snowcover-free season, the 7-week shortening of the snowcover-free season produced by the combination of orbital, CO₂ and modest tundra forcing in Exp-C might have been sufficient to produce widespread incipient glaciation poleward of 60° N. Without this bias, it is possible that even a smaller tundra expansion than was used in Exp-C might have been sufficient to induce incipient glaciation over areas of northeastern Canada.

Further investigation of orbitally induced glaciation will necessitate considerable improvement in representation of atmospheric physics in GCMs (better determination of clouds, snow accumulation and snow melt) and coupling to the terrestrial biosphere. The models should be of higher resolution than used here in order to represent better the weather processes producing precipitation and snow and to resolve the detailed land–sea configuration and topography for the areas likely to have been sites of glaciation. Increased model resolution will allow the more realistic depiction of local regions of tundra expansion rather than the broad expansion used here. Future work should include coupling of atmospheric models with ocean GCMs to examine the role of feedbacks involving the thermohaline circulation^{3,4} on the initiation of glaciation, and the potential interplay of such feedbacks with land-surface processes. □

Received 17 November 1995; accepted 30 April 1996.

- Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
- Imbrie, J. & Imbrie, K. P. *Science* **207**, 943–953 (1980).
- Imbrie, J. et al. *Paleoceanography* **7**, 701–738 (1992).
- Imbrie, J. et al. *Paleoceanography* **8**, 699–735 (1993).
- Rind, D., Petet, D. & Kukla, G. J. *geophys. Res.* **94**, 12851–12871 (1989).
- Royer, J. F., Déqué, M. & Pestiaux, P. in *Milankovitch and Climate* (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 733–764 (Reidel, Hingham, MA, 1984).
- Galimore, R. G. & Kutzbach, J. E. *J. geophys. Res.* **100**, 1103–1120 (1995).
- Phillips, P. J. & Held, I. M. *J. Climate* **7**, 767–782 (1994).
- Mitchell, J. F. B. *Phil. Trans. R. Soc. Lond. B* **341**, 267–275 (1993).
- Dong, B. & Valdes, P. J. *J. Climate* **8**, 2471–2496 (1995).
- Harrison, S. P., Kutzbach, J. E., Prentice, I. C., Behling, P. J. & Sykes, M. T. *Quat. Res.* **43**, 174–184 (1995).
- Bonan, G. B., Pollard, D. & Thompson, S. L. *Nature* **359**, 716–718 (1992).
- Foley, J., Kutzbach, J. E., Coe, M. T. & Lewis, S. *Nature* **371**, 52–54 (1994).
- Berger, A., Gallée, H. & Tricot, C. *J. Glaciol.* **39**, 45–49 (1993).
- Covey, C. & Thompson, S. L. *Paleogeogr. Paleoclimatol. Paleoeconol.* **76**, 331–341 (1989).
- Barnola, J.-M., Piment, P., Raynaud, D. & Korotkevich, Y. S. *Tellus* **43B**, 83–90 (1991).
- Ramanathan, V., Lian, M. S. & Cess, R. D. *J. geophys. Res.* **84**, 4949–4958 (1979).
- Oglesby, R. J. & Saltzman, B. *J. Climate* **5**, 66–92 (1992).
- Huntley, B. & Webb, T., III. (eds) *Vegetation History* (Kluwer, Dordrecht, 1988).
- Pons, A., Guiot, J., de Beaulieu, J. L. & Reille, M. *Quat. Sci. Rev.* **11**, 439–448 (1992).
- de Vernal, A., Miller, G. H. & Hillaire-Marcel, C. *Quat. Int.* **10–12**, 95–106 (1991).
- Robinson, D. A. & Kukla, G. J. *Clim. appl. Met.* **24**, 402–411 (1985).
- Boulton, G. S. & Clark, C. D. *Trans. R. Soc. Edinb.* **82**, 327–347 (1990).
- Climate Mapping and Prediction (CLIMAP) Series MC-36 (Geological Society of America, 1981).
- Peltier, W. R. *Science* **265**, 195–201 (1994).
- Berger, A. et al. *Trans. R. Soc. Edinb.* **82**, 357–369 (1990).
- Chappell, J. & Shackleton, N. J. *Nature* **324**, 137–140 (1986).
- Bromwich, D. H., Tzeng, R.-Y. & Parish, T. R. *J. Climate* **7**, 1051–1069 (1994).

ACKNOWLEDGEMENTS. We thank P. Behling and R. Selin for preparing the figures and M. Kennedy for technical preparation of the manuscript; D. Robinson and P. Bartlein for their assistance with information about observational studies. This research was sponsored by U.S. National Science Foundation grants to the University of Wisconsin-Madison. The computer calculations were performed at the National Center for Atmospheric Research, which is supported by the US NSF.

The mechanism of dissolution of oxide minerals

William H. Casey* & Christian Ludwig†

* Department of Land, Air and Water Resources, and Department of Geology, University of California, Davis, California 95616, USA

† Swiss Institute for Environmental Science and Technology (EAWAG), CH-8600 Dübendorf, Switzerland.

THE exchange of metal ions between an oxide mineral surface and water occurs in a wide range of processes, including corrosion¹, the breakdown of inhaled dusts^{2,3}, soil formation⁴ and the cycling of toxic substances in the environment⁵. In studies of the mechanisms of dissolution, the measured rate-law order with respect to protons^{6–15} cannot be reconciled with the number of protons needed to form any reasonable assumed activated complex. Here we suggest that this discrepancy can be avoided if one takes into account the number of protonation and deprotonation steps leading to detachment of the hydrated metal ion. We show that the experimental proton rate order reflects a net balance of protons removed and attached in these steps. Our mechanism

explains why the rate order generally coincides with the metal valence^{8,9,11,12,16–18}, and why there is a similarity between rates of water ligand lability in dissolved complexes and rates of mineral dissolution^{19–22} and metal desorption²³. It eliminates the need to invoke catalysis by protons, and establishes a close consistency between reactions at surfaces and (better understood) ligand-exchange reactions in solution.

Experimental rate orders are derived by fitting adsorbed ligand and proton concentrations to a steady-state rate law. For dissolution at conditions where positive surface charge predominates and far from equilibrium, the rate is proportional to the concentration of adsorbed protons at the surface raised to some order n (refs 8–13): rate = $k_{\text{H}^+} x_{\text{SOH}_2^+}^n$, where k_{H^+} is a rate coefficient and $x_{\text{SOH}_2^+}$ is the proton adsorbate concentration. The rate coefficient yields the rate in terms of $\text{mol cm}^{-2} \text{s}^{-1}$ and the rate order, n is the key parameter used to interpret the stoichiometry of proposed activated complexes and is presumed to indicate the number of protons configured to a detaching metal in an activated complex (see refs 9–12 and 15 for example). Some research groups^{8,9,11,12,17} report that n equals the metal valence in the solid but this result is not general^{6,7,13,14}. The experiments, however, are difficult and results, particularly for mixed-oxide solids, are sometimes ambiguous. Furthermore, none of these models accounts for the observation^{19–22} that rates of metal removal from surfaces, including

TABLE 1 Changes in the coordination of surface oxygens on monoatomic steps as metal sites are protonated and detached as $\text{Fe}(\text{H}_2\text{O})_6^{3+}$, $\text{Ni}(\text{H}_2\text{O})_6^{2+}$ and $\text{Al}(\text{H}_2\text{O})_6^{3+}$.

[001] Goethite*									
Pair no. 1		Pair no. 2		Pair no. 3		Pair no. 4		... Pair no. n	
Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	... 1 → 0	1 → 0
1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	... 1 → 0	1 → 0
2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	... 2 → 1	2 → 1
2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	... 2 → 1	2 → 1
3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	... 3 → 2	3 → 2
3' → 2	3' → 2	3' → 2	3' → 2	3' → 2	3' → 2	3' → 2	3' → 2	... 3' → 2	3' → 2
New site exposures									
		$\bar{3} \rightarrow 3$	$\bar{3} \rightarrow 3$	$\bar{3} \rightarrow 3$	$\bar{3} \rightarrow 3$	$\bar{3} \rightarrow 3$	$\bar{3} \rightarrow 3$	... $\bar{3} \rightarrow 3$	$\bar{3} \rightarrow 3$
		$\bar{3}' \rightarrow 3'$	$\bar{3}' \rightarrow 3'$	$\bar{3}' \rightarrow 3'$	$\bar{3}' \rightarrow 3'$	$\bar{3}' \rightarrow 3'$	$\bar{3}' \rightarrow 3'$	... $\bar{3}' \rightarrow 3'$	$\bar{3}' \rightarrow 3'$
[111] NiO(s)†									
No. 1		No. 2		No. 3		... No. 4			
1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	... 1 → 0			
2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	... 2 → 1			
3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	... 3 → 2			
4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	... 4 → 3			
5 → 4	5 → 4	5 → 4	5 → 4	5 → 4	5 → 4	... 5 → 4			
6 → 5	6 → 5	6 → 5	6 → 5	6 → 5	6 → 5	... 6 → 5			
New site exposures									
		$\bar{6} \rightarrow 6$	$\bar{6} \rightarrow 6$	$\bar{6} \rightarrow 6$	$\bar{6} \rightarrow 6$	... $\bar{6} \rightarrow 6$			
[001] $\alpha\text{-Al}_2\text{O}_3$ ‡									
A		B		C		D		E	
1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0
1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0	1 → 0
2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1	2 → 1
2 → 1	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2	3 → 2
2 → 1	3 → 2	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3
4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3	4 → 3
New site exposures									
		$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$
		$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$
		$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$	$\bar{4} \rightarrow 4$

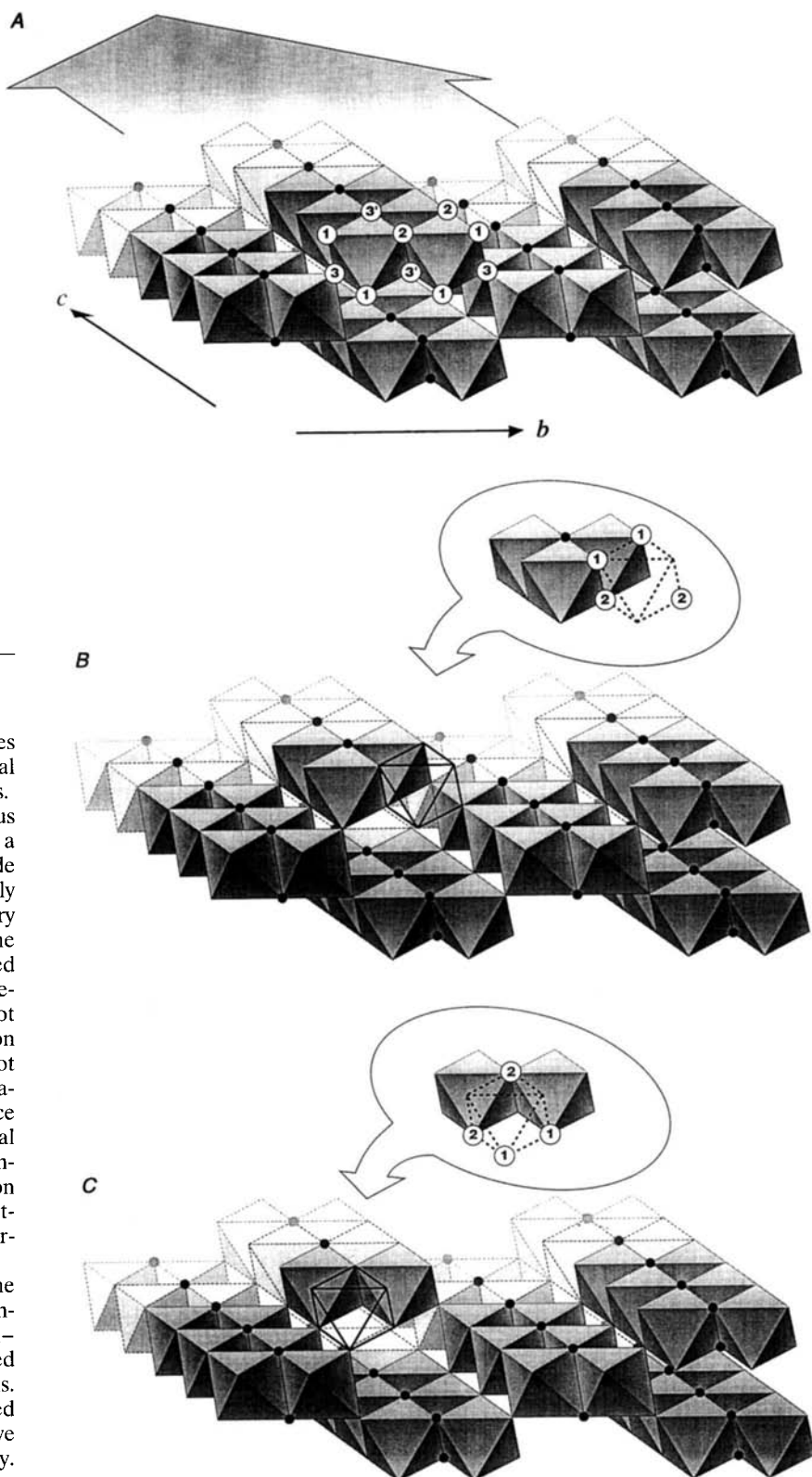
The first one or two octahedra for all step orientations are coordinatively unique because they are adjacent to the edge of the crystal. Steady state is achieved after these octahedra are detached.

* Fig. 1.

† Fig. 2.

‡ Fig. 3.

FIG. 1 Double-chains of $\text{Fe(III)}\text{O}_6^{2-}$ octahedra exposed at the [001] surface of goethite $[\text{FeO}(\text{OH})\text{(s)}]$. Solid symbols (●) represent structural hydroxyl groups. Steady-state dissolution involves sequential protonation and removal of FeO_6^{2-} octahedra from double chains parallel to the *c*-axis of the mineral. These chains retreat as metals detach to form what is essentially a monoatomic step. The labelling scheme (A) identifies the number of iron atoms coordinated to each oxygen and are keyed to the reactions in Table 1. The labels 1 and 3, for example, indicate oxygens coordinated to one and three iron atoms, respectively. The oxygen labelled 3' identifies exposed structural hydroxyl sites. Detachment (B, C) of an $\text{Fe}(\text{H}_2\text{O})_6^{3+}$ eliminates six total oxygens, of which two are dangling sites ($1 \rightarrow 0$) and four must be re-established at the surface with reduced coordination numbers. Two of these sites change $2 \rightarrow 1$ and two change $3 \rightarrow 2$ (Table 1). The wire frame (B, C) identifies the detaching octahedra. The inset (B, C), with the detaching octahedron now dashed, shows the reduced coordination numbers of oxygens on surrounding octahedra after re-establishment of the surface through movement and dissociation of water molecules.



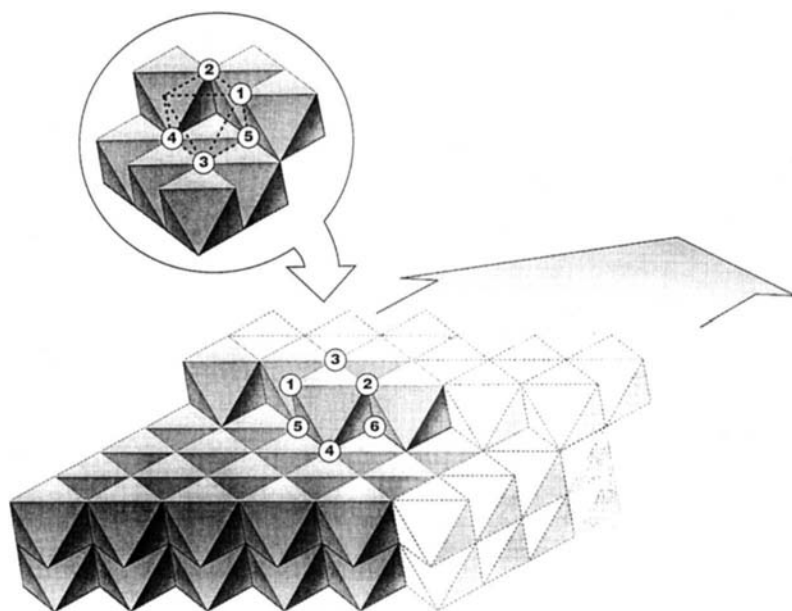
metal–ligand complexes^{21,24}, behave like the rates of water exchange between the bulk and the metal hydration sphere in similar dissolved complexes.

The concept of a steady-state surface allows us to reinterpret the rate order and to postulate a mechanism. Protons adsorb so rapidly to oxide surfaces that an equilibrium state is usually achieved¹¹. Conventional transition-state theory states that the rate order corresponds to the number of protons needed to form the activated complex and that protonation reactions subsequent to decomposition of this complex are not included in the rate law. As long as protonation occurs at the dissolving surface, however, and not on the detached complex in solution, it is immaterial whether the protons adsorb to the surface before or after detachment of a particular metal site. Protonation reactions that follow one detachment precede the next (it is the rate of metal ion detachment, not of protonation, that is rate limiting; detachments from monoatomic steps are particularly important^{25,26}).

Goethite provides a useful example of the dissolution process because the structure is complicated by structural hydrogens. The metal–oxygen octahedra (Fig. 1) are linked by shared edges to form double chains parallel to the *c*-axis. At any exposed surface, oxygens are coordinated to three, two or one Fe(III) atoms, which we identify as 3 sites, 2 sites and 1 sites, respectively. The 3' sites are exposed structural hydroxyl sites, which are identified for counting purposes but undergo rapid acid–base reactions at the surface in the same manner as 3 sites. Bulk structural oxygens that are not exposed at the surface are labelled with an overbar ($\bar{3}$ and $\bar{3}'$), and oxygens coordinated to a metal in solution are denoted 0 sites. The reaction $1 \rightarrow 0$ denotes loss of a dangling oxygen from the surface as the surface complex detaches to form a solute; the reaction $\bar{3} \rightarrow 3$ indicates that a three-coordinated oxygen is uncovered that was previously buried in the bulk structure.

Proton rate orders are most often measured in acid solutions^{7–21,27} where the fully hydrated metal is most stable, so we choose $\text{Fe}(\text{H}_2\text{O})_6^{3+}$ as the stoichiometry of the detaching surface complex. The number of protons needed to form this complex is counted by protonating oxygens in the FeO_6^{2-} octahedra to form water molecules. The coordinated oxygens need two protons apiece to form water molecules so that the total number of

FIG. 2 Dissolution of a monoatomic step from the [111] surface of a divalent metal oxide mineral with the rocksalt structure, such as NiO(s). The structure is cubic so the step lies perpendicular to either the [100] or [010] face. In the main figure, numbers identify different coordination numbers for oxygens on a MO_6^{10-} octahedra on the surface. A 6 site, for example, represents the oxygen coordinated to six metals and that lies at a common corner for six octahedra. Removal of each divalent metal ($\text{Ni}(\text{H}_2\text{O})_6^{2+}$ for example) eliminates six oxygens, one of which is a dangling oxygen [$1 \rightarrow 0$]. Five of the oxygens are re-established on the surfaces with decreased metal coordination numbers [$6 \rightarrow 5$, $5 \rightarrow 4$, $3 \rightarrow 2$, $2 \rightarrow 1$, $1 \rightarrow 0$] and a new six-coordinated site is exposed [$\bar{6} \rightarrow 6$] upon removal of each octahedrally coordinated metal. The inset, with the detaching octahedron now dashed, shows the reduced coordination numbers of oxygen on surrounding octahedra after the surface is re-established through movement and dissociation of water molecules.



protons is 12 if there are no protons on the oxygens before reaction. Adjustment is made for pre-existing protons on the oxygens by setting x , y and z equal to the number of protons coordinated to 3 (and 3'), 2 and 1 oxygen sites, respectively. For goethite, there are two 3 sites, two 2 sites and two 1 sites surrounding each detaching $\text{Fe}(\text{III})$ (Table 1). The total number of protons needed to form $\text{Fe}(\text{H}_2\text{O})_6^{3+}$ at the surface is $12 - 2x - 2y - 2z$.

Removal of each $\text{Fe}(\text{H}_2\text{O})_6^{3+}$ from the steady-state surface (Fig. 1) eliminates two dangling oxygens [$2 \cdot (1 \rightarrow 0)$], but four of the removed oxygens must be replaced to fulfil the coordination sphere of surface iron atoms. These missing oxygens are replaced through reaction with water molecules, which dissociate to yield the equilibrium concentration of protons at each site (y or z). It is at present impossible to tell whether the detaching or remaining iron atoms get oxygens from the solvent, but such distinction is not needed to understand the mechanism.

Once re-established, the four oxygens at the surface have reduced coordination numbers to iron atoms [$3' \rightarrow 2$, $3 \rightarrow 2$, $2 \rightarrow 1$, $2 \rightarrow 1$], so that the net number of protons released as the equilibrium protonation states are re-established is $(2 \cdot 4) - 2y - 2z$. Steady state requires that two 3 and 3' sites uncover as each $\text{Fe}(\text{H}_2\text{O})_6^{3+}$ detaches and these sites are exposed at the surface of the underlying chain. Protonation of the 3 site to achieve an equilibrium state consumes x protons and protonation of the 3' site, a structural hydroxyl, consumed $x - 1$ protons.

The net number of protons consumed by detaching a single $\text{Fe}(\text{H}_2\text{O})_6^{3+}$ from the monoatomic step is $n = (12 - 2x - 2y - 2z) - (8 - 2y - 2z) + (2x - 1) = 3$, the experimental result^{11,12}. Note that n provides no information about the equilibrium protonation states of surface oxygens around a detaching metal; x , y and z can have any value including zero. The rate order, n , is determined solely by the difference between the number of protons consumed by detachment of a surface metal complex and the number of protons released by movement and dissociation of water. A similar result would be found for any crystallographically accurate surface as long as consecutive reactions yield similar sets of oxygens and the fully hydrated metal is chosen as the detaching species. For example, steady-state dissolution of monoatomic steps on the [111] face of a rocksalt oxide (Fig. 2) yields the following changes in oxygen coordination numbers: [$\bar{6} \rightarrow 6$, $6 \rightarrow 5$, $5 \rightarrow 4$, $3 \rightarrow 2$, $2 \rightarrow 1$, $1 \rightarrow 0$]. Again, rate order equals the metal valence, the experimental value^{8,17,18}. $n = (12 - u - v - w - x - y - z) - (2 \cdot 5 - v - w - x - y - z) + (u) = 2$, but not the number of adsorbed protons in a precursor complex.

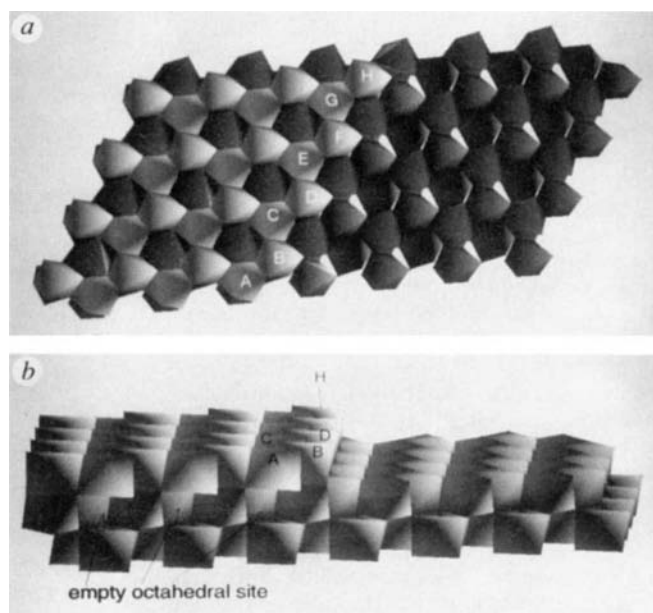


FIG. 3 a, View onto the [001] surface of corundum [$\alpha\text{-Al}_2\text{O}_3$] with a monoatomic step lying perpendicular to the [100] orientation. The light-coloured polyhedra lie on the surface as a monoatomic step and individual AlO_6 polyhedra on this step are identified by letters A–H that key to Table 1. The structure consists of AlO_6 octahedra, each of which has a shared face parallel to the [001] plane. Polyhedra also share edges and corners, so that each oxygen in the bulk is coordinated to four metals, and one third of the octahedral sites are unoccupied. Every AlO_6 polyhedron in the monoatomic step overlies unoccupied octahedral sites (A, C, E, G in this figure). As these octahedra detach, they expose three new four-coordinated oxygens to the surface [$3 \cdot (4 \rightarrow 4)$]. b, Side view of the monoatomic step, enabling the unoccupied octahedral sites to be most easily seen.

Similarly, n is independent of the equilibrium protonation states (values of $u - z$) of surface oxygens.

Dissociation of a monoatomic step on the [001] surface of $\alpha\text{-Al}_2\text{O}_3$ to release $\text{Al}(\text{H}_2\text{O})_6^{3+}$ (Fig. 3) provides a more complicated example. This structure includes face-, edge- and corner shared AlO_6 octahedra and steady-state is achieved by removing pairs of octahedra because every second octahedron, labelled A, C, E, ..., overlies three four-coordinated ($4 \rightarrow 4$) oxygens. The number

of protons needed to detach two $\text{Al}(\text{H}_2\text{O})_6^{3+}$ species is: $24 - 3w - 3x - 3y - 3z$, where w , x , y and z are the number of protons at equilibrium on four-, three-, two- and single-coordinated oxygens. There are three of each oxygen type [4, 3, 2, 1] in the reaction, and not two as in the case of goethite (Table 1). The newly exposed bulk sites consume $3w$ protons to achieve equilibrium at the surface and the nine vacated coordination sites react with water molecules and release $(18 - 3x - 3y - 3z)$ protons to achieve their equilibrium protonation states. The net number of protons consumed by steady-state reaction is: $6 = (24 - 3w - 3x - 3y - 3z) - (18 - 3x - 3y - 3z) + 3w$ or three protons per aluminium; $n = 3$, as observed in some^{8,17} but not all¹³ experiments.

Each monoatomic step on these minerals is essentially a polymer of identical linked metal-oxide polyhedra. The $\text{M}(\text{H}_2\text{O})_6^{2+}$ monomer detaching from these steps has 12 protons, but the net number of protons transferred from the aqueous solution to each metal before detachment is much smaller than 12. In these examples, it equals the metal valence and the proton rate order, and reflects the stoichiometry on the detaching complex. If, for example, $\text{Fe}(\text{H}_2\text{O})_5(\text{OH})^{2+}$ detaches from the goethite surface in acid solutions, $n = 2$ not $n = 3$. Other cases where n does not equal the metal valence probably indicate that the experimental surfaces are not at a steady state²⁷; data from experiments on mixed-oxide minerals are complicated by cation leaching and should not be fully trusted^{27,28}.

Although here we focus on proton reactions, this dissolution mechanism is general and requires a fundamental reassessment of the existing rate laws and reaction models. By assigning ligand functional groups to metal-coordination sites, one can easily demonstrate that ligand-promoted dissolution is rarely independent of adsorbed protons. In addition, rates of reductive dissolution can be organized into classes where either electron transfer or metal detachment is slow. For the latter class, rate orders and reactivities can be predicted in the same way as here, because movement of water molecules from the bulk solution to the inner-coordination sphere of a metal controls the reaction rates. $\square$

Emplacement of the Taupo ignimbrite by a dilute turbulent flow

W. Brian Dade & Herbert E. Huppert

Institute of Theoretical Geophysics, Department of Earth Sciences and Department of Applied Mathematics & Theoretical Physics, University of Cambridge, Cambridge CB2 3EQ, UK

An ignimbrite is a pumice-rich deposit that records the passage of a ground-hugging ash flow (a 'pyroclastic flow') generated by the collapse of a volcanic eruption column¹. Geologists study such deposits to reconstruct the parent eruptions and to predict the consequences of future eruptions. The 1,800-yr-old Taupo ignimbrite of New Zealand has been interpreted to represent *en masse* emplacement by an avalanche-like flow with a volumetric solids concentration in excess of 30% (ref. 2). The evidence for this is equivocal, however, and here we propose the alternative view that the deposit was emplaced by a relatively dilute and turbulent density current³. We present an isothermal, hydraulic model, the results of which, taken together with existing observations, suggest that the total flow rate of solids and gas was about $40 \text{ km}^3 \text{ s}^{-1}$ for around 15 minutes. This intense flux resulted in a flow which had a near-vent solids concentration of 0.3% by volume, was about 1 km thick, and travelled outward from the vent with a typical speed of 200 m s^{-1} . In view of the good agreement between predicted and observed radial trends in the Taupo deposit, we suggest that the origin of other ignimbrites with similar characteristics should be reconsidered.

The Taupo eruption of 1.8 kyr ago was one of the largest explosive eruptions of the past 10,000 years (ref. 4), and Wilson's detailed account² of the Taupo ignimbrite makes it the best-documented deposit of its kind. The ignimbrite comprises about 30 km^3 of material spread over $20,000 \text{ km}^2$ of terrain with low regional slope, and is axisymmetrically distributed around the inferred vent. Owing to its widespread nature, the Taupo deposit is universally regarded as a 'low-aspect-ratio ignimbrite', which resulted from an extremely explosive eruption⁵. The mechanism of emplacement of this and similar deposits, however, is currently one of the most contentious issues in volcanology⁶, and features of the Taupo deposit often cited by field geologists as evidence for a densely concentrated parent flow are the subject of continuing debate. For example, intense particle segregation associated with ingestion of ambient air overrun by the flow front, documented by the presence of a relatively coarse, bipartite basal layer ('layer 1'), is not unique to a densely concentrated granular flow. On the contrary, it is only for turbulent, dilute gravity currents that significant entrainment of ambient fluid into the region of the flow front has been well established⁷⁻⁹. Similarly, gas-escape and particle-segregation structures in the main body of the deposit, previously taken to reflect the importance of fluidization as a support mechanism in the parent flow, are now recognized as potential products of late-stage depositional processes unrelated to the dynamics of the regional transport system^{10,11}.

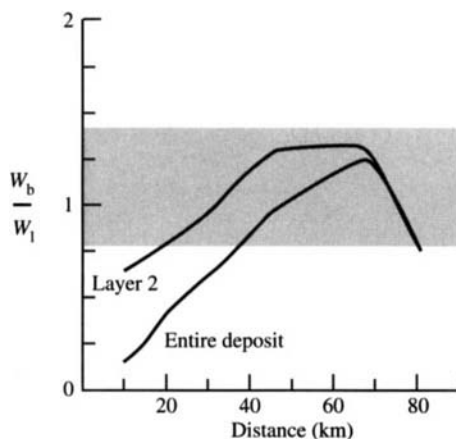
The fact that extremely large, individual pumices with diameters of up to 60 cm were transported to distances of 45 km from the vent has been among the strongest arguments invoked against a turbulent parent flow^{2,12}. A dilute, turbulent flow should produce a deposit in which particles are (approximately) aerodynamically equivalent, that is, one in which particles with similar settling velocities are found together in similar proportions. Thus it has been argued that the far-travelled pumice boulders could only have 'floated' to their point of rest on a parent flow in which the density of the matrix of fines exceeded that of individual pumice boulders. This reasoning is based on incorrect notions of settling of particles with individual diameters in excess of about

Received 12 February; accepted 24 April 1996.

1. Blesa, M. A., Morando, P. J. & Regazzoni, A. E. *Chemical Dissolution of Metal Oxides* (CRC, Boca Raton, FL, 1994).
2. Kuhn, D. C. & Demers, L. M. J. *Toxicology and Environmental Health* **35**, 39–50 (1992).
3. Hochella, M. F. in *Health Effects of Mineral Dusts* (eds Guthrie, D. G. & Mossman, B. T.) 275–305 (Rev. Mineral. **28**, Mineral. Soc. Am., 1993).
4. Hilgard, E. W. *Soils: Their Formation, Properties, Compositions and Relations to Climate and Plant Growth in the Humid and Arid Regions* 19 (MacMillan, New York, 1914).
5. Davis, J. A. & Leckie, J. O. in *Chemical Modeling in Aqueous Systems* (ed. Jenne, E. A.) (Am. chem. Soc., Washington DC, 1979).
6. Wirth, G. S. & Gieskes, J. M. J. *Colloid Interface Sci.* **68**, 492–500 (1979).
7. Blum, A. & Lasaga, A. C. *Geochim. cosmochim. Acta* **55**, 2193–2201 (1991).
8. Furrer, G. & Stumm, W. *Geochim. cosmochim. Acta* **50**, 1847–1860 (1986).
9. Wieland, E., Wehrli, B. & Stumm, W. *Geochim. cosmochim. Acta* **52**, 1969–1981 (1988).
10. Ganor, J., Mogollón, J. L. & Lasaga, A. C. *Geochim. cosmochim. Acta* **59**, 1037–1053 (1995).
11. Stumm, W. *Chemistry of the Solid–Water Interface* (Wiley, New York, 1992).
12. Stumm, W. & Wollast, R. *Rev. Geophys.* **28**, 53–69 (1990).
13. Carroll-Webb, S. A. & Walther, J. V. *Geochim. cosmochim. Acta* **52**, 2609–2623 (1988).
14. Puffer, K., Schindler, P. W., Westall, J. C. & Grauer, R. J. *Colloid Interface Sci.* **101**, 554–564 (1984).
15. Oelkers, E. H. & Schott, J. *Geochim. cosmochim. Acta* **59**, 5039–5054 (1995).
16. Guy, C. & Schott, J. *Chem. Geol.* **78**, 181–204 (1989).
17. Ludwig, C. & Casey, W. H. J. *Colloid Interface Sci.* **178**, 176–185 (1996).
18. Furrer, G. thesis, ETH Zurich (1985).
19. Casey, W. H. J. *Colloid Interface Sci.* **146**, 586–589 (1991).
20. Casey, W. H. & Westrich, H. R. *Nature* **355**, 157–159 (1992).
21. Ludwig, C., Casey, W. H. & Rock, P. A. *Nature* **375**, 44–47 (1995).
22. Dove, P. M. & Czanek, C. A. *Geochim. cosmochim. Acta* **59**, 1907–1916 (1995).
23. Hachiya, K., Sasaki, M., Ikeda, T., Mikami, N. & Yasunaga, T. *J. phys. Chem.* **88**, 27–31 (1984).
24. Ludwig, C., Devidal, J.-L. & Casey, W. H. *Geochim. cosmochim. Acta* **60**, 213–224 (1996).
25. Gratz, A. J., Manne, S. & Hansma, P. K. *Science* **251**, 1343–1346 (1991).
26. Hochella, M. F. in *Mineral–Water Interface Geochemistry* (eds Hochella, M. F. & White, A. F.) Ch. 3 (Rev. Mineral. **23**, Mineral. Soc. Am., Washington DC, 1990).
27. Casey, W. H., Westrich, H. R., Banfield, J. F., Feruzzi, G. & Arnold, G. *Nature* **366**, 253–256 (1993).
28. Casey, W. H. & Bunker, B. in *Mineral–Water Interface Geochemistry* (eds Hochella, M. F. & White, A. F.) Ch. 9 (Rev. Mineral. **23**, Mineral. Soc. Am., Washington DC, 1990).

ACKNOWLEDGEMENTS. We thank Drs. G. Sposito, W. Stumm, P. Rock, B. Phillips, J. Banfield and H. Westrich for helpful discussions and R. Higashi for the photograph of Fig. 3. Support was from the US NSF and DOE.

CORRESPONDENCE should be addressed to W.H.C. (e-mail: whcasey@uodavis.edu) or C.L. (e-mail: christian.ludwig@eawag.ch).



0.1 mm in low-density gases^{13–15}. The settling velocity of such particles is approximately $(g\rho'd)^{1/2}$, where g is the acceleration due to gravity, ρ' is the ratio of particle-to-gas phase densities, and d is nominal diameter of the particle. Given that the density of boulders of large pumice ρ_b is 200–500 kg m⁻³ and that of lithic fragments ρ_l is 2,400–2,700 kg m⁻³ (ref. 2), the ratio of the diameter of the largest pumices to that of the largest lithics should be 5–14 for the largest particles of either type to be in aerodynamic equivalence. Alternatively, the ratio of settling velocities $W_b/W_l \approx [(\rho'd)_b/(\rho'd)_l]^{1/2}$ should be approximately unity at all transport distances. This is the case (Fig. 1). The presence of pumice boulders in the Taupo ignimbrite at large transport distances is entirely consistent with emplacement by a dilute, turbulent flow.

Additional confirmation that the deposit was emplaced by such a flow comes from the degree to which individual grains are sorted by size in the transport direction (Fig. 2b, c). These patterns are not generally seen in deposits emplaced by debris avalanches¹⁶, a commonly invoked analogue for densely concentrated pyroclastic flows. Distal fining is common, on the other hand, in deposits emplaced by dilute, turbulent gravity currents in other settings¹⁷. In particular, the combination of massive bedding with little or no vertical grading at a given locality and well defined lateral sorting indicates that rapid deposition took place from a vertically well-mixed suspension of particles in a horizontally non-uniform, turbulent flow subject to constant flux and sediment supply for the limited duration of the event¹⁸.

To evaluate quantitatively the concentration of the Taupo flow, we report here the results of a general hydraulic model for a deposit-forming gravity current¹⁹, modified to accommodate a range of grain sizes in the driving suspension. The model describes the runout of a turbulent gravity current subject to a constant flux at the vent and which spreads over a landscape of small regional slope. It incorporates quantitative expressions for the evolution of flow volume, a Froude-number condition at the flow front, and the gravitational settling of particles from the driving suspension in the body of the flow (see Box 1).

The effects of entrainment of ambient air, time-dependent thermodynamics and compressibility are not considered in describing the deposit-forming flow, and we assume that the driving suspension is sufficiently dilute that interactions between suspended particles are negligible. The model works well even when the density contrast between the ambient fluid and the gravity current is large²⁰, and in particular for subaqueous deposit-forming currents with solids concentrations of up to 30% by volume²¹. Thus by 'dilute' we mean volumetric solids concentrations less than 30%. We suggest that this approach includes the essential physical processes that govern the runout behaviour of a wide class of deposit-forming gravity currents. In geological terms, the model describes the regional transport system that gives rise to

FIG. 1 Ratio of settling velocities of largest pumice boulders (W_b) and lithic fragments (W_l) with nominal densities of 350 kg m⁻³ and 2,500 kg m⁻³, respectively, as a function of radial distance from the vent. Settling velocities are calculated with the analytical approximation used in our model (see Box 1) for maximal sizes reported by Wilson² for layer 2 and for the entire deposit. The shaded region indicates the range of values that correspond to approximate aerodynamic equivalence, which accommodates the observed range of densities of respective particle types. The low values near the vent probably reflect source-controlled scarcity of large pumices overall and the presence of reworked, pre-existing lithics in layer 1.

BOX 1 Equations describing a radially spreading gravity current

We describe the propagation of a deposit-forming current by solving the governing axisymmetric equations for a flow in which the total flux is constant¹⁹. These have solutions that agree well with numerical treatment of the full equations²⁸. The current radius $R(t)$, thickness $h(t)$, speed $U(t)$ and solids concentration $C(t)$ are related to the total flux Q and the settling velocity W by

$$Q = 2\pi R h U, \quad U = dR/dt = Fr[g'_p(C - C_*)h]^{1/2}, \quad \frac{dC}{dt} = -WC/h \quad (1)$$

where $Fr \approx 1$ is the Froude number at the head of the current propagating over relatively flat terrain¹⁹, and $g'_p = (\rho_p - \rho_i)g/\rho_a$ is the reduced gravity of the current expressed in terms of g , the acceleration due to gravity, and ρ_p , ρ_i and ρ_a , the densities of individual particles, interstitial gas and ambient air, respectively. The term $C_* = (\rho_a - \rho_i)/(\rho_p - \rho_i)$ represents the concentration of suspended solids at which the flow becomes positively buoyant. The equations given in (1) represent, respectively, conservation of total flux, hydraulic control at the head and the loss of particles through sedimentation.

The settling velocity W of a spheroidal particle with nominal diameter d falling through an interstitial gas with kinematic viscosity ν is

$$W = \nu/d[(81 + \eta_*)^{1/2} - 9] \quad (2)$$

where $\eta_* = g'_p \rho_a d^3 / \rho_i \nu^2$. Equation (2) is derived from the balance between gravitational and drag forces which govern the settling of individual particles and for which the drag coefficient is $1 + 24\nu/Wd$. It is a useful approximation for all particle sizes.

When modified to accommodate the probability distribution $p(\phi)$ of particle settling speeds in the initial suspension in terms of the logarithmically transformed particle size ϕ , the total solids concentration in the flow at radial distance R is

$$C = C_0 \int_{-\infty}^{\infty} \exp[-\pi W(\phi)R^2/Q] p(\phi) d\phi \quad (3)$$

where C_0 is constrained to lie between C_* and a value required by conservation of sedimentary mass.

The deposit thickness is given by

$$D = D_0 \int_{-\infty}^{\infty} W(\phi) \exp[-\pi W(\phi)R^2/Q] p(\phi) d\phi / \int_{-\infty}^{\infty} W(\phi) p(\phi) d\phi \quad (4)$$

where D_0 is the observed thickness of the deposit near the vent. The composition of the deposit at a distance R comes directly from that of the overlying suspension. Hence the relative amount of material C_d equal to or finer than size ϕ_d that is preserved in the deposit at R is given by

$$C_d = \int_{\phi_d}^{\infty} \exp[-\pi W(\phi)R^2/Q] p(\phi) d\phi / \int_{-\infty}^{\infty} \exp[-\pi W(\phi)R^2/Q] p(\phi) d\phi \quad (5)$$

The calculations shown in Fig. 2 are for an eruption column and flow with constant temperature 450 °C and with interstitial gas in which air and water vapour are present in equal amounts. Ambient air is 10 °C. Hence ρ_i , ν , and ρ_a are 0.4 kg m⁻³, 0.8 cm² s⁻¹ and 1.2 kg m⁻³, respectively. We assume an average value ρ_p of 1,000 kg m⁻³ and, following Wilson's observations², that particle sizes in the source eruption column were (approximately) uniformly distributed in the size range $(7\phi) - (-7\phi)$.

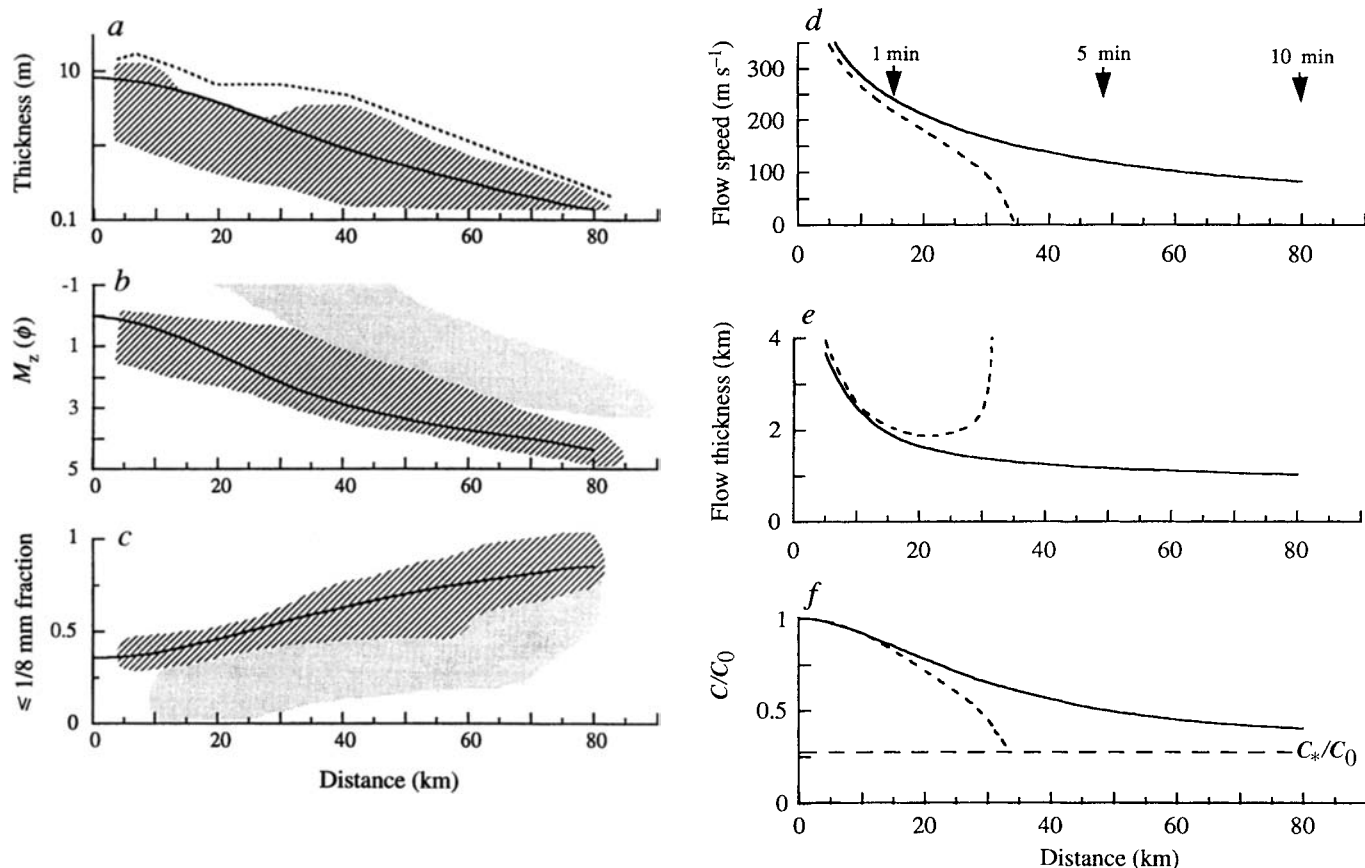


FIG. 2 Characteristics of the Taupo ignimbrite and the parent flow. Radial trends in deposit characteristics as follows. *a*, Deposit thickness. *b*, Graphic mean grain size M_z in units of ϕ , where $\phi = -\log_2$ (diameter in mm) (see ref. 2). *c*, Fraction of material smaller than or equal to 1/8 mm (3ϕ). Characteristics of a reconstructed parent flow for which the temperature and particle concentration in the eruption column were respectively 450°C and 3 kg m^{-3} . *d*, Flow speed. *e*, Flow thickness. *f*, Concentration of solids relative to concentration at the vent. The shaded regions in *a*–*c* indicate the range of existing observations² for the veneer deposit found throughout the Taupo region. The grain-size data do not include the pumice concentration zones and so are biased to finer (larger ϕ) sizes. The dashed

line in *a* shows the observed total thickness if maximal estimates for the thickness of layer 1 are included. The lightly stippled regions in *b* and *c* indicate lateral sorting trends in the relatively coarse layer 1 deposits. The solid lines in *a*–*f* show model calculations for the polydisperse suspension observed within the bulk deposit. The short-dashed lines in *d*–*f* represent model calculations for a suspension characterised by a single fall speed of particles (4 m s^{-1}) corresponding to that for the median grain size of 1 mm. The horizontal dashed line in *f* indicates the relative concentration $C_*/C_0 = 0.27$ ($\sim 1\text{ kg}$ of solids per m^3) at which the model flow would have become neutrally buoyant.

large-scale, downstream variations in a turbulent flow and its deposit, rather than the depositional system that results in features in the deposit observed on outcrop scale^{22,23}.

Inputs to the model include material properties for solid and fluid phases. The results are relatively insensitive to exact values of these properties if they are within environmentally imposed constraints. Here we assume that the average density of individual particles is $1,000\text{ kg m}^{-3}$, that the temperature of the eruption column and flow was 450°C and that air and water in the interstitial gas were present in equal amounts. Beyond the general assumption of a dilute state, we do not assume *a priori* a specific value of initial solids concentration in the parent flow. Rather, we calculate the concentration of solids using our model and the constraints imposed upon it by the deposit. There are no freely adjustable parameters in the model.

With the incorporation of the known thickness of accumulation near the vent and the bulk grain-size characteristics of the entire deposit (ref. 2), our model predicts radial trends in the thickness and sorting of grain sizes which are in good agreement with Wilson's observations for the veneer deposit (Fig. 2*a*–*c*). An important exception is that the sediments in layer 1 are generally coarser than the overlying veneer deposit as a result, perhaps, of the intense processes of sorting and elutriation associated with the mixing of ambient air into the flow front. The veneer deposit, in

combination with the granulometrically similar ponded facies, nevertheless comprises a significant fraction of the overall volume of the ignimbrite and reflects deposition from the main body of the parent flow. Accordingly, it is this depositional unit which is most indicative of the regional transport system in the parent flow. Our model describes the process of particle sorting in the downstream direction based on differences in the settling speeds of individual grains. The good agreement shown in Fig. 2 between predicted and observed measures of deposit thickness (representing a weighted accumulation of all grain sizes), mean grain size and the distribution of fines implies that all grain sizes in the ignimbrite are in approximate aerodynamic equilibrium (as are the largest pumices and lithic fragments discussed above).

Using our model, we infer from the deposit that the total volumetric flux in the turbulent parent flow was of the order of $40\text{ km}^3\text{ s}^{-1}$. No additional parameters are required for the model predictions for the deposit characteristics shown in Fig. 2*a*–*c*. This simplicity constitutes a potentially powerful result of our analysis. We also note, however, that conservation of mass during a finite period of constant flux constrains the concentration of solids in the collapsing eruption column to have been about 3 kg m^{-3} . This degree of expansion means that the associated ash flow was of the order of 1 km thick and advanced at a speed initially in excess of 250 m s^{-1} . It reached the 80-km extent of the known deposit in no

more than about 15 minutes, by which time the eruption apparently ceased and the ash flow dissipated. Our reconstruction of the deposit-forming flow is presented in Fig. 2d–f. These show, among other things, that the extended runout of the flow beyond about 40 km was due to the presence of fine particles in the parent suspension.

The total flux and near-vent solids concentration inferred for the Taupo event represent values which resulted from the collapse of the eruption column combined with continued discharge of new material at the vent. These values accordingly impose an upper limit on the rate of magma eruption of $5 \times 10^7 \text{ m}^3 \text{ s}^{-1}$. Given our assumptions regarding temperature and relative vapour content, this suggests that ejecta in the eruption column were at approximately 700 °C and that the fraction of water vapour was about 5% by weight. This result is consistent with independent estimates of water content in the eruption column²⁴. Such conditions correspond to a scenario in which material erupted at vertical speeds in excess of 300 m s⁻¹ from a vent with a radius of almost 1 km into a collapsing, fountain-like column²⁵. The extensive dilution of vented material in the column, and thus its mobility in the resulting turbulent pyroclastic current, reflect the extreme explosivity of the eruption. Preliminary results of detailed calculations indicate that subtle changes in the inferred initial properties of the parent flow do not alter the basic picture presented here.

We are confident that additional details of the Taupo deposit can be reconciled with more sophisticated models for turbulent-flow emplacement. Some near-vent features of the ignimbrite, for instance, may be the result of inhomogeneities in particle con-

centration in the parent eruption column²⁶. Abrupt changes in fines and lithic contents in the ponded deposits at distances of 50–60 km from the vent, on the other hand, may reflect changes in the degree of stratification of the transporting flow^{15,22} or interaction of the flow with regional topography which, on average at these distances, comprises ridge-like obstacles that are 500–1,000 m above vent level. These details are qualitatively consistent with our approach but are not explicitly accommodated.

The ease with which our quantitative model can predict regional patterns in deposit characteristics at Taupo argues strongly for the reconsideration of other ignimbrites with similar geometry and sorting trends previously interpreted solely in terms of the prevailing model of emplacement by a highly concentrated flow. Broad consistency of field data with a turbulent-flow model alone does not, of course, exclude other mechanisms of emplacement. In our view, an important challenge is to determine additional deposit properties which can be used to differentiate emplacement mechanisms and relate these to eruption history and associated volcanic hazards. A potentially powerful tool in this regard is the measurement of anisotropy of magnetic susceptibility²⁷ which reveals the fabric of iron-bearing, elongate grains in a deposit which, in turn, is related to the mechanism of emplacement. Given the current developments in both theory and observation, we suggest that the term 'low-aspect-ratio ignimbrite' be applied only to the physical characteristics of a deposit, and be stripped of connotations regarding the mechanism of emplacement. □

Received 13 September 1995; accepted 22 April 1996.

1. Sparks, R. S. J., Self, S. & Walker, G. P. L. *Geology* **1**, 115–118 (1973).
2. Wilson, C. J. N. *Phil. Trans. R. Soc. Lond. A* **314**, 229–310 (1985).
3. Fisher, R. V. *Am. J. Sci.* **264**, 287–298 (1966).
4. Wilson, C. J. N. & Walker, G. P. L. *Phil. Trans. R. Soc. Lond. A* **314**, 199–228 (1985).
5. Walker, G. P. L., Heming, R. F. & Wilson, C. J. N. *Nature* **283**, 286–287 (1980).
6. Valetine, G. A. & Fisher, R. V. *Science* **259**, 1130–1131 (1993).
7. Simpson, J. E. *Acta mechanica* **63**, 245–253 (1986).
8. Hallworth, M. A., Phillips, J. C., Huppert, H. E. & Sparks, R. S. J. *Nature* **362**, 829–831 (1993).
9. Hallworth, M. A., Huppert, H. E., Phillips, J. C. & Sparks, R. S. J. *J. Fluid Mech.* **308**, 289–311 (1996).
10. Branney, M. J. & Kokelaar, P. *Bull. volcan.* **54**, 504–520 (1992).
11. Druitt, T. H. J. *Volcan. geotherm. Res.* **65**, 27–39 (1995).
12. Wilson, C. J. N. *Bull. volcan.* **50**, 350–351 (1988).
13. Batchelor, G. K. *An Introduction to Fluid Mechanics* (Cambridge Univ. Press, 1983).
14. Clift, R., Grace, J. R. & Weber, M. E. *Bubbles, Drops, and Particles* (Academic, London, 1978).
15. Valentine, G. A. *Bull. volcan.* **50**, 352–355 (1988).

16. Voight, B. *Rocksides and Avalanches*, *Dev. Geotech. Eng.* Vol. 14A (Elsevier, Amsterdam, 1978).
17. Pickering, K. T., Hiscott, R. N. & Hein, F. J. *Deep Marine Facies* (Unwin Hyman, London, 1989).
18. Kneller, B. C. & Branney, M. J. *Sedimentology* **42**, 607–616 (1995).
19. Dade, W. B. & Huppert, H. E. *J. geophys. Res.* **100**, 18597–18609 (1995).
20. Hopfinger, E. J. A. *Rev. Fluid Mech.* **15**, 47–76 (1983).
21. Dade, W. B. & Huppert, H. E. *Geology* **22**, 645–648 (1994).
22. Valentine, G. A. *Bull. volcan.* **49**, 616–630 (1987).
23. Fisher, R. V. *Geol. Soc. Am. Bull.* **102**, 1038–1054 (1990).
24. Dunbar, N. W. & Kyle, P. R. *J. Volcan. geotherm. Res.* **49**, 127–145 (1992).
25. Sparks, R. S. J., Wilson, L. & Hulme, G. J. *geophys. Res.* **83**, 1727–1739 (1978).
26. Anilkumar, A. V., Sparks, R. S. J. & Sturtevant, B. J. *Volcan. geotherm. Res.* **56**, 145–160 (1993).
27. Fisher, R. V. *J. Volcan. geotherm. Res.* **56**, 205–220 (1993).
28. Bonnetaze, R. T., Huppert, H. E. & Lister, J. R. *J. Fluid Mech.* **294**, 93–121 (1995).

ACKNOWLEDGEMENTS. R. Cas, T. Druitt, D. Pyle, S. Sparks, G. Valentine, C. Wilson and A. Woods provided constructive and challenging discussions of pyroclastic flows and ignimbrites, although not all agree with the views presented here.

Eusociality in a coral-reef shrimp

J. Emmett Duffy

School of Marine Science and Virginia Institute of Marine Science,
The College of William and Mary, Gloucester Point, Virginia 23062, USA

THE apex of animal social organization is eusociality, which has three characteristics: overlapping generations, reproductive division of labour, and cooperative care of young^{1,2}. So far, eusociality has been recognized only among social insects and the African mole-rats^{3–5}. Here I report the first case of eusociality in a marine animal. The sponge-dwelling shrimp *Synalpheus regalis* lives in colonies that may have >300 individuals, but that contain only one reproductive female. Direct-developing juveniles remain in the natal sponge, and allozyme data suggest that most colony members are full sibs. In laboratory experiments, larger colony members, most of whom apparently never breed, defended the colony against heterospecific intruders. Ecological similarities among mole-rats, termites and these sponge-dwelling shrimp, all of which are diploid animals, strengthen arguments that eusociality is favoured by gradual

metamorphosis, parental care, and occupation of protected, expandable niches⁶.

The mobile cryptofauna of coral reefs is frequently dominated by snapping shrimps (*Synalpheus*)^{7–9}, many of which live within the internal canals of sponges. *S. regalis* is an abundant cryptofaunal species at Carrie Bow Cay, Belize, and inhibits the sponges *Xestospongia* cf. *subtriangularis* and *Hyattella intestinalis*^{9,10}. Each of over 30 dissected sponges contained a colony of *S. regalis* consisting of 3–313 individuals ($\bar{x}$ = 149, from 17 complete colonies examined). As tropical *Synalpheus* species apparently breed continuously (J.E.D., personal observation), reproductive females are easily identified by ripe ovaries or brooded eggs, yet each colony invariably contained only a single reproductive female. The remaining juveniles and mature males are difficult to distinguish morphologically, probably reflecting a socially mediated sex-determination system, as is known in another alpheid¹¹. By having just one female, these colonies clearly meet the first criterion of reproductive division of labour.

Eusociality also requires that generations overlap², bringing individuals into contact with close kin such that altruistic behaviours can be favoured by kin selection¹². Three lines of evidence indicate that *S. regalis* colonies consist primarily of sibling offspring of the lone female (the 'queen'). First, *S. regalis* eggs hatch into crawling juveniles; successive cohorts of small juveniles are clearly recognizable in many colonies, suggesting that they nor-

mally remain in the natal sponge. Second, larger (presumably older) queens have accumulated larger colonies than have smaller ones; queen body size explains 65%, and fecundity 75%, of the variation in colony size (Fig. 1). Finally genotype frequencies at three polymorphic allozyme loci yield an average value for r (relatedness¹³) of 0.57 (Table 1), close to the value of 0.50 expected for full sibs. This suggests that most colony members (there may be >300) are offspring of the queen, and possibly of a single male. If indeed a single 'king' fathers most offspring (the possibility of inbreeding complicates this conclusion), he is not obviously distinguished by morphology.

The distinctive feature of eusociality is the altruistic behaviour of non-breeders, particularly cooperative feeding and defence of young. Sponge-dwelling shrimps apparently obtain food from the host's feeding current and/or tissues^{10,14,15}, such that food within the sponge is probably not limiting. Instead, the biggest challenge facing shrimp colonies in the coral-rubble habitat, where unoccupied habitable sponges are virtually non-existent^{8,10}, is probably invasion of the nest by competitors. Responses of colonies to conspecific intruders have not yet been tested, but the role of non-breeders in defending the colony against interspecific competitors was demonstrated by exposing colonies in the laboratory to heterospecific intruders and to previously isolated colony-mates ('natives') (Fig. 2). The results were dramatic: contact between a resident and a 'foreign' intruder (an undescribed species near *Synalpheus bousfieldi*) generally resulted in an intense battle, with both individuals repeatedly snapping at one another with their powerful major chelae (claws). This continued sporadically until the intruder was killed. In two of ten experimental runs, two residents together gripped the vanquished intruder's carcass and dragged it out of the chamber. In stark contrast, contacts between residents and 'natives' were quite peaceful. Importantly, larger residents were most active and aggressive, contacting foreign

intruders more than twice as often as did smaller residents, and engaging intruders in combat (snapping) ten times more often than did juveniles (Fig. 2). Because larger individuals, most of whom are unlikely ever to reproduce, perform most of the colony defence, and thereby benefit juveniles, this nest defence amounts to cooperative brood care, fulfilling the third criterion of eusociality.

The finding of eusociality in a diploid marine shrimp is especially intriguing as *S. regalis* exhibits precisely those characteristics hypothesized to have fostered eusociality in termites and mole-rats, the two previously known diploid, eusocial taxa, namely 'gradual metamorphosis, subsociality (extensive parental care), and life in long-lasting, expansible niches (nests or microhabitats) safe from predation and rich with food that does not require exiting the safety of the niche to obtain it' (ref. 6, p. 42). The advantage of cooperation in defending such nests is widely believed to be a primary selective pressure for advanced social life^{2,5,6,16-20}, and cavity nesting is indeed characteristic of nearly all eusocial taxa, including most ants²¹, all termites², mole-rats⁵, ambrosia beetles²², and gall-dwelling aphids²³ and thrips²⁴. Similarly, living within sponge canals appears to have been an important preadaptation for eusociality in *S. regalis*, as many of its

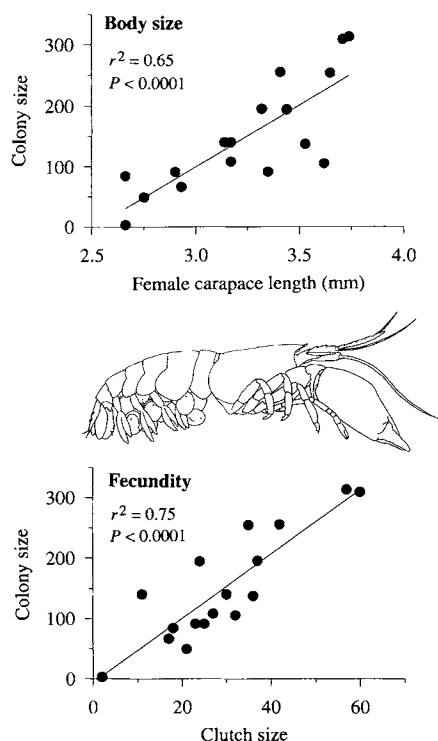


FIG. 1 Positive relationship between the body size and fecundity of the queen, and total colony size in *S. regalis*. Complete shrimp colonies were dissected from 17 sponges, *Xestospongia* cf. *subtriangularis* and *Hyattella intestinalis*, from the outer reef ridge (15–20 m depth) at Carrie Bow Cay, Belize; all individuals were counted, the carapace length (an index of body size) measured, and the number of embryos carried by the female in each colony counted.

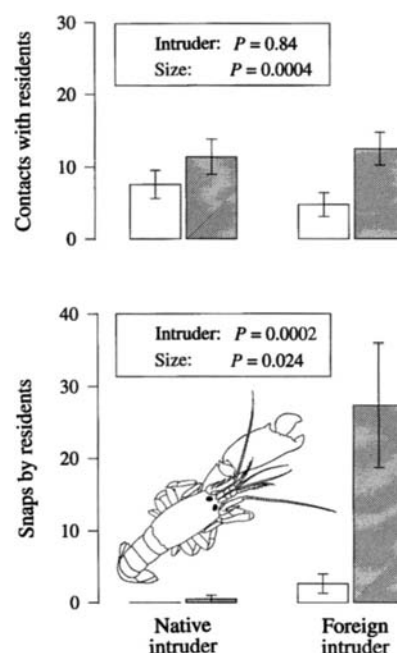


FIG. 2 Responses of laboratory colonies of *Synalpheus regalis* to intrusion by a heterospecific competitor, versus intrusion by their own colony-mates. Experiments were conducted in viewing chambers consisting of a system of canals cut in neoprene and pressed between transparent plastic plates. After submerging the assembly in a flow-through aquarium, part of a colony of shrimp (the female, 8 large males, and 8 juveniles) was introduced and allowed to acclimate overnight. In the morning, two experimental trials were conducted sequentially: (1) a member of the resident colony ('native') that had been isolated overnight was introduced to the chamber; and (2) a 'foreign' intruder of an undescribed species near *S. bousfieldi* was introduced. The order of introduction was alternated in successive trials ($n = 8$ trials). The colony was observed for 30 min, during which all contacts and snaps between residents and intruder were recorded, along with the size (small, white bars; large, shaded bars) of the resident involved in these interactions. Foreign intruders that remained alive at the end of 30 min were invariably killed by residents within a few hours. P -values were obtained by randomly reassigning the observed counts among the four treatment cells within a 'block' (that is, the replicates introducing foreigner and native to the same colony of residents) 10,000 times and calculating the proportion of these randomizations in which differences as great as or greater than the observed values were found. The P -value for Intruder is for the difference between responses to native versus foreign intruders, irrespective of size class; similarly, the P -value for Size is for the difference between small and large residents, irrespective of intruder type. Bars show $\bar{x} \pm 1$ s.e.m.

TABLE 1 Estimates of average genetic relatedness among colony members in *Synalpheus regalis*

Locus	<i>n</i>	Average <i>r</i>
<i>Pgi</i>	208	0.50 (0.14)
<i>Tpi</i>	201	0.66 (0.13)
<i>Pep</i>	194	0.56 (0.15)
Mean	208	0.57 (0.04)

Genotypes were scored, using standard methods⁹, at three polymorphic allozyme loci in 17–22 individuals from each of 10 colonies (*n* is the total number of individuals used in a calculation) collected along a ~200-m transect on the outer reef ridge at Carrie Bow Cay, Belize. Relatedness (*r*) was estimated according to the expression given in ref. 13, with s.e. (in parentheses) estimated by jack-knifing over colonies; the estimates were obtained using the program Relatedness 4.2 (K. F. Goodnight). The mean value of within-colony relatedness, 0.57, is quite close to that (0.50) expected for full sibs, and suggests that colony members are offspring of a single breeding pair.

sponge-dwelling congeners live in groups of tens to hundreds per sponge^{8,14,15,23}, and direct development is common among such species^{26,27} (J.E.D., unpublished observation). In contrast, most other alpheidids produce planktonic larvae, live in heterosexual pairs, and are aggressive towards other conspecifics^{11,26–27}. Eusociality in *Synalpheus* may also have been fostered by possession of a powerful weapon, the major chela, as argued for the sting of Hymenoptera and perhaps the conspicuous weapons of other eusocial taxa^{2,20,23,24}. Interestingly, the likelihood of eusociality among crustaceans has been predicted³⁰, based in part on the direct development and cavity-dwelling habits of many crustaceans. Indeed, I have collected two other monogynous, probably eusocial, species of *Synalpheus* in Belize. This discovery supports conclusions that eusociality in disparate animal taxa was favoured by kin selection interacting with strong natural selection imposed by enemies. □

Received 26 January; accepted 23 April 1996.

1. Michener, C. D. A. *Rev. Ent.* **14**, 229–342 (1969).
2. Wilson, E. O. *The Insect Societies* (Belknap, Harvard Univ. Press, Cambridge, MA, 1971).
3. Jarvis, J. U. M. *Science* **212**, 571–573 (1981).
4. Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D. *The Biology of the Naked Mole-Rat* (Princeton Univ. Press, NJ, 1991).
5. Jarvis, J. U. M., O'Riain, M. J., Bennett, N. C. & Sherman, P. W. *Trends Ecol. Evol.* **9**, 47–51 (1994).
6. Alexander, R. D., Noonan, K. M. & Crespi, B. J. in *The Biology of the Naked Mole-Rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 3–44 (Princeton Univ. Press, NJ, 1991).
7. Snelgrove, P. V. R. & Lewis, J. B. *Mar. Biol.* **101**, 249–257 (1989).
8. Duffy, J. E. *Mar. Ecol. Prog. Ser.* **90**, 127–138 (1992).
9. Duffy, J. E. *Biol. J. Linn. Soc. Lond.* (in the press).
10. Duffy, J. E. *J. crust. Biol.* (in the press).
11. Nakashima, Y. *J. Ethol.* **5**, 145–159 (1987).
12. Hamilton, W. D. J. *theor. Biol.* **7**, 1–52 (1964).
13. Queller, D. C. & Goodnight, K. F. *Evolution* **43**, 258–275 (1989).
14. Rützler, K. *Tethys* **7**, 249–264 (1976).
15. Erdman, R. B. & Blake, N. J. *J. crust. Biol.* **7**, 328–337 (1987).
16. Lin, N. & Michener, C. D. *Q. Rev. Biol.* **47**, 131–159 (1972).
17. Michener, C. D. *Proc. 10th Int. Congr. Ent., Montreal, 1956* Vol. 2, 442–447 (1958).
18. Wilson, E. O. *Sociobiology: The New Synthesis* (Belknap, Harvard Univ. Press, Cambridge, MA, 1975).
19. Evans, H. E. *BioScience* **27**, 613–617 (1977).
20. Andersson, M. A. *Rev. Ecol. Syst.* **15**, 165–189 (1984).
21. Hölldobler, B. & Wilson, E. O. *The Ants* (Belknap, Harvard Univ. Press, Cambridge, MA, 1990).
22. Kent, D. S. & Simpson, J. A. *Naturwissenschaften* **79**, 86–87 (1992).
23. Itô, Y. *Trends Ecol. Evol.* **4**, 69–73 (1989).
24. Crespi, B. J. *Nature* **359**, 724–726 (1992).
25. Dardeau, M. R. *Mem. Hourglass Cruises* **7**(2), 1–125 (1984).
26. Knowlton, R. E. *J. nat. Hist.* **7**, 273–306 (1973).
27. Dobkin, S. R. *F.A.O. Fish. Rep.* **57**, 935–946 (1969).
28. Knowlton, N. *Evolution* **34**, 161–173 (1980).
29. Knowlton, N. & Keller, B. D. *Bull. mar. Sci.* **37**, 893–904 (1985).
30. Spanier, E., Cobb, S. J. & James, M.-J. *Oikos* **67**, 573–576 (1993).

ACKNOWLEDGEMENTS. I thank K. Rützler and M. Carpenter for facilitating work at Carrie Bow Cay; K. Rützler and K. Smith for confirming sponge identifications; R. Grosberg and J. Graves for hospitality and discussion; C. Morrison for assistance; M. Sidall for the randomization analysis of the behavioural data; and E. A. Canuel, S. F. Craig, R. K. Grosberg, J. L. King, N. Knowlton, P. B. Marko and E. O. Wilson for comments on the manuscript. This work was supported by the National Science Foundation (DEB postdoctoral fellowship) and the Smithsonian Institution's Caribbean Coral Reef Ecosystems program.

CORRESPONDENCE and requests for materials should be addressed to the author (e-mail: jeduffy@vims.edu).

Embryology of vestimentiferan tube worms from deep-sea methane/sulphide seeps

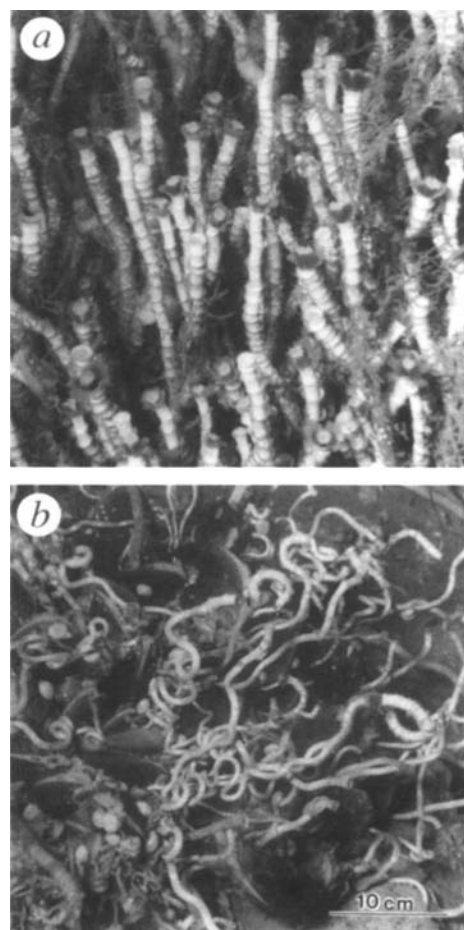
Craig M. Young*, Elsa Vázquez*, Anna Metaxas* & Paul A. Tyler†

* Department of Larval Ecology, Harbor Branch Oceanographic Institution, 5600 U.S. Highway 1 North, Fort Pierce, Florida 34946, USA

† Department of Oceanography, University of Southampton, Southampton SO9 5NH, UK

THE Vestimentifera are gutless worms that live around deep-sea hydrothermal vents and cold seeps, obtaining energy from hydrogen sulphide with the aid of endosymbiotic chemosynthetic bacteria^{1–3}. Their phylogenetic relationships have been debated ever since they were first discovered^{4,5}. Moreover, hydrothermal vents are ephemeral and spatially patchy, raising questions about how vestimentiferan populations are established and maintained^{6–9}, and how symbionts are transmitted¹⁰. Although post-settling juveniles have been described^{11,12}, embryos and larvae have been neither collected nor cultured. Here we describe the early development of vestimentiferans from cold seeps in the Gulf of Mexico¹³, and discuss the implications of our findings for dispersal potential and phylogeny.

Lamellibrachia sp. and *Escarpia* sp. (Fig. 1) were collected by manned submersible and dissected immediately. Details of the culture methods are given in the figure legends; developmental descriptions apply to both species unless stated. Eggs are optically dense and positively buoyant, those of *Escarpia* sp. being slightly



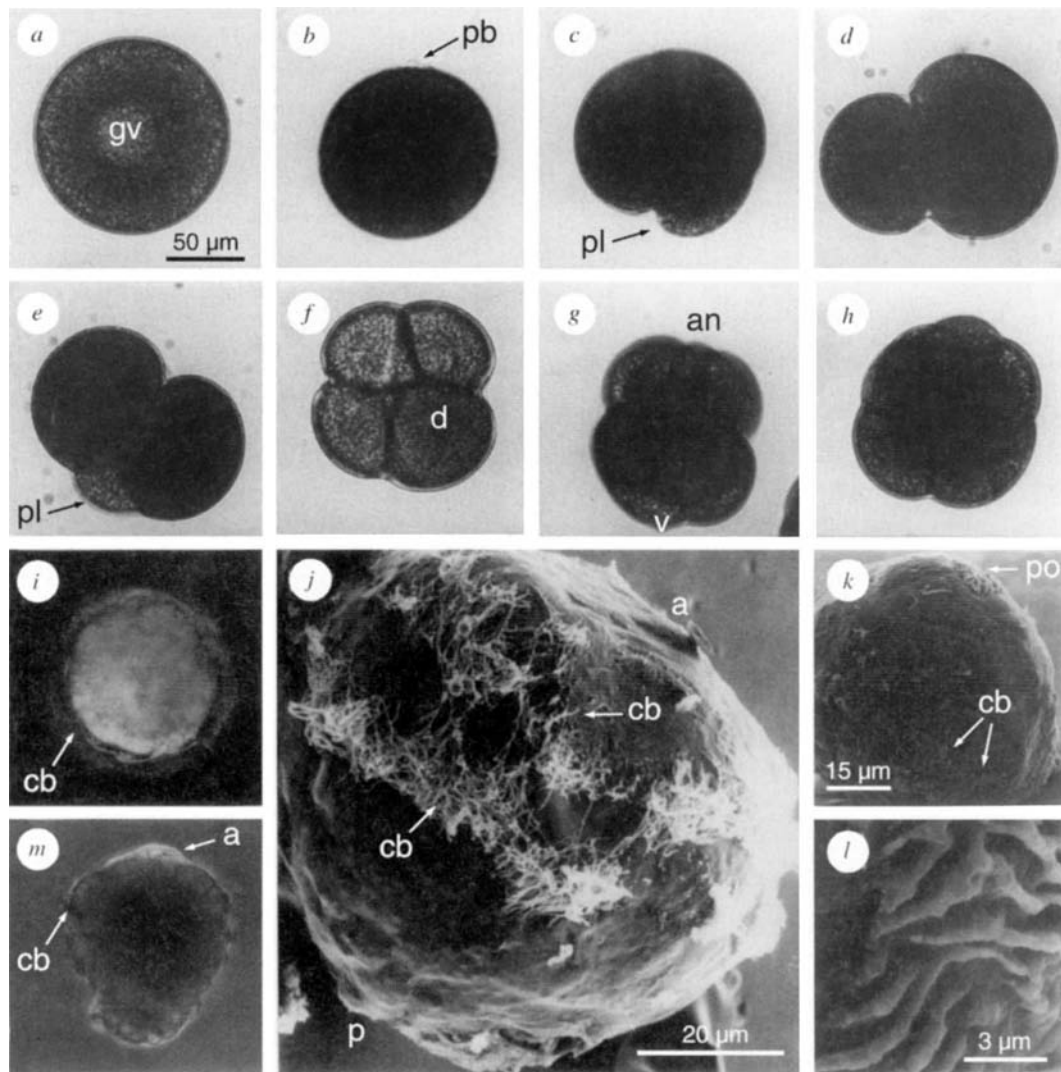


FIG. 2 Embryos and larvae of vestimentiferans. Mature gametes were dissected from the anterior portion of the gonoduct, located midventrally near the posterior margin of the vestimentum. Gametes were inseminated *in vitro* and the embryos were cultured in 0.45 µm filtered seawater at 1 atm, 37 parts per thousand salinity, and 9 °C with water changes every second day. Embryos reared in pressure vessels at 50 and 100 atm developed at the same rate as embryos in control cultures at 1 atm. Light micrographs were taken on the ship with a Nikon Labophot microscope. a, Immature ovum of *Escarpia* sp. with germinal vesicle (gv) on dissection from the oviduct. The scale bar also applies to b–i and m. b, Embryo of *Lamellibrachia* sp. 3-h old, showing absence of an obvious germinal vesicle and presence of two polar bodies (pb), the large first polar body on the left, and the small second one on the right. c, Embryo of *Escarpia* sp. just before first cleavage, showing first polar lobe (pl). d, Two-cell embryo of *Lamellibrachia* sp. showing unequal blastomere size. e, Formation of

second polar lobe (pl) in *Lamellibrachia* sp. just before second cleavage. f, Four-cell embryo of *Escarpia* sp. showing large 'D' cell (d). g, Lateral view of an eight-cell embryo of *Escarpia* sp. Note that the upper tier of blastomeres (marked 'an' for animal pole) is shifted 45 degrees relative to the vegetal tier (v) as in classic spiral cleavage. h, Anterior view of an eight-cell embryo of *Escarpia* sp., showing spiral cleavage. i, Larva (12 days old) of *Escarpia* sp., viewed from the anterior end. Note the strong equatorial cilia (cb). j, SEM of a 10-day-old larva of *Escarpia* sp. showing presence of two ciliary bands; a, anterior end (note absence of apical tuft); p, posterior end. Specimens were fixed for SEM by sudden immersion in 4% OsO₄ in Millonigs phosphate buffer. k, Lateral view of a 10-day-old *Escarpia* sp. larva, showing position of posterior organ (po) in relation to the ciliary band. l, Enlarged view of the posterior organ of *Escarpia* sp. larva. m, Larva (9 days old) of *Lamellibrachia* sp., which is beginning to elongate posteriorly.

FIG. 1 a, A dense stand of *Lamellibrachia* sp. showing many animals with obturacula and lamellae extended. This species is abundant near methane seeps, where tubes up to 2 m long occur in massive bushes. Adult tube worms were collected with a Johnson-Sea-Link submersible during September 1995 from Bush Hill (27°46.949'N, 91°30.490' W) and other sites in the Green Canyon Block of the Louisiana Slope, all about 600 m in

depth. b, Highly coiled tubes of *Escarpia* sp. The tubes may be up to 1 m long, and are generally found near the periphery of *Lamellibrachia* bushes or in other areas of hydrocarbon seepage, as indicated by bacterial mats and mussels (*Bathymodiolus* sp.), as shown here. The scale bar applies to both a and b.

larger (115 μm , s.d. = 3.67, $N = 25$) than those of *Lamellibrachia* sp. (105.55 μm , s.d. = 3.67, $N = 25$). Liberated eggs had large germinal vesicles (Fig. 2a) and completed maturation divisions within 3 hours (Fig. 2b). Sperm were bundled within the gonoducts, but individual sperm broke apart and swam upon dilution. Cleavage began 14 hours after insemination (Fig. 2c), and subsequent divisions occurred at intervals of approximately 4 hours. Formation of a large polar lobe (Fig. 2c) resulted in unequal first cleavage along the animal/vegetal axis (Fig. 2d). A second polar lobe formed before the second cleavage (Fig. 2e), yielding a large 'D' cell and three smaller blastomeres of about equal size (Fig. 2f). Third cleavage was equatorial and spiral (Fig. 2g, h). The eight-cell embryo resembled that of a typical polychaete more closely than that of a brooding perviate pogonophore with elongate ova¹⁴⁻¹⁶.

Larvae became ciliated 3 days after fertilization. They remained at the surface until they were 8 days old, when they began to swim down through the water column. The young larvae of *Escarpi* sp. were nearly spherical and trochophore-like, with two bands of apparently simple cilia placed equatorially (Fig. 2i-k); by contrast, most polychaete trochophores have compound cilia^{17,18}. Scanning electron microscopy revealed no mouth, apical tuft, or telotroch. Simple cilia are less efficient than compound cilia at capturing food¹⁸, but may not be essential if larvae remain lecithotrophic throughout larval life. The posterior end of the larva had a circular region of plications of unknown function (Fig. 2k, l). Larvae of the two species were virtually identical until about 9 days, when those of *Lamellibrachia* began to elongate (Fig. 2m). By 21 days, the pretrochal regions of both species were only about one-third the length of their respective posttrochal regions. No mouth or gut was visible even at this stage.

Vestimentiferans are regarded as close relatives of the perviate Pogonophora¹¹, which were initially placed among the oligomeric deuterostome phyla¹⁶ but are now considered to be proto-stomes. Opinion is divided about whether these two groups constitute a discrete phylum⁵ or whether they are highly specialized annelids¹⁹, as supported by our findings, as well as recent genetic analyses²⁰. Early vestimentiferan embryology is nearly identical to that of oviparous polychaetes, having similar egg size^{21,22}, common features include spiral cleavage, two polar lobes²¹, unequal cleavage resulting in a large 'D' cell in the four-cell stage²¹, and a trochophore-like larva with two parallel bands of cilia. Juveniles of *Ridgeia* found in the tubes of adults are also very similar to annelids, having complete guts¹¹, setae²³ and proto-trochs¹². The spiralian nature of early embryology makes it unlikely that enterocoelous mesoderm formation will be found (as has been reported for perviate pogonophorans²⁴). Indeed, vestimentiferan embryology seems more similar to that of polychaetes than to that of perviates^{14,15}, although this contrast may be artificial, as all known perviate embryos are brooded in narrow tubes.

Despite major interest in the colonization of vents and seeps⁶⁻¹⁰, no previous study has given any estimate of the length of vestimentiferan larval life. Our study shows that embryos and larvae float up into the water column, where they disperse as lecithotrophs for at least several weeks. We do not know if larvae eventually require planktonic food or if they ever become capable of feeding. Whatever the mode of larval nutrition, dispersal potential of seep vestimentiferans is enormous.

Note added in proof: Recent taxonomic work indicates that *Escarpi* sp. probably belongs to an undescribed genus (S. Gardiner, personal communication). □

Received 2 February; accepted 25 March 1996.

1. Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. & Waterbury, J. B. *Science* **213**, 340-342 (1981).
2. Felbeck, H., Powell, M. A., Hand, S. C. & Somero, G. N. *Bull. Biol. Soc. Wash.* **6**, 261-272 (1985).
3. Childress, J. J., Arp, A. J. & Fisher, C. R. *Mar. Biol.* **83**, 109-124 (1984).
4. Webb, M. *Bull. mar. Sci.* **19**, 18-47 (1969).
5. Jones, M. L. *Bull. Biol. Soc. Wash.* **6**, 117-158 (1985).
6. Lutz, R. A., Jablonski, D. & Turner, R. D. *Science* **226**, 1451-1454 (1984).
7. Kim, S. L., Mullineaux, L. S. & Helfrich, K. R. *J. geophys. Res.* **99** (C6), 12655-12665 (1994).

8. Mullineaux, L. S., Wiebe, P. H. & Baker, E. T. *Oceanus* **34**, 64-68 (1991).
9. Smith, C. R., Kukert, H., Wheatcroft, R. A., Jumars, P. A. & Deming, J. W. *Nature* **341**, 27-28 (1989).
10. Cary, S. C., Felbeck, H. & Holland, N. D. *Mar. Ecol. Prog. Ser.* **52**, 89-94 (1989).
11. Southward, E. C. *J. mar. biol. Ass. U.K.* **68**, 465-487 (1988).
12. Jones, M. L. & Gardiner, S. L. *Biol. Bull.* **177**, 254-276 (1989).
13. MacDonald, I. R. et al. *Mar. Biol.* **101**, 235-247 (1989).
14. Southward, E. C. in *Reproduction of Marine Invertebrates Vol. II* (eds Giese, A. C. & Pearse, J. S.) 129-156 (Academic, New York, 1975).
15. Bakke, T. *Sarsia* **60**, 1-11 (1976).
16. Ivanov, A. V. *Pogonophora* (Academic, New York, 1963).
17. Nielsen, C. *Acta zool.* **68**, 205-262 (1987).
18. Emlet, R. B. & Strathmann, R. R. in *Reproduction and Development of Marine Invertebrates* (eds Wilson, W. H., Stricker, S. A. & Shinn, G. L.) 143-157 (Johns Hopkins Univ. Press, Baltimore, 1994).
19. Land, J. Van der & Nørrevang, A. Z. *zool. Syst. Evolutionsforsch.* **1975**, 86-101 (1975).
20. Kojima, S. et al. *J. molec. Evol.* **37**, 66-70 (1993).
21. Schroeder, P. C. & Hermans, C. O. in *Reproduction of Marine Invertebrates Vol. III* (eds Giese, A. C. & Pearse, J. S.) 1-213 (Academic, New York, 1975).
22. Anderson, D. T. *Acta zool.* **47**, 1-42 (1966).
23. Bartolomeus, T. *Zoomorphology* **115**, 161-177 (1995).
24. Gureeva, M. A. & Ivanov, A. V. *Zool. Zh.* **65**, 780-788 (1986).

ACKNOWLEDGEMENTS. We thank the captain and crew of the RV *Edwin Link*, and the support crew of the JSL II submersible; K. Eckelbarger, M. Devin, J. Zardus, M. Baker, L. Hawkins, and T. Ward were present on the cruise and helped with many aspects of this work; M. Rice, M. Sewell, C. Cary and C. Fisher participated in useful discussions; M. Rice and E. Ruppert read a draft manuscript; and J. Piraino assisted with SEM work. This work was supported by the National Undersea Research Program of NOAA through the University of North Carolina at Wilmington.

CORRESPONDENCE and requests for materials should be addressed to C.M.Y. (e-mail: youngc@hboi.edu).

Role of the thrombin receptor in development and evidence for a second receptor

Andrew J. Connolly*†‡, Hiroaki Ishihara*†, Mark L. Kahn*†§, Robert V. Farese Jr||§ & Shaun R. Coughlin*†§

* Cardiovascular Research Institute, † Daiichi Research Center, Departments of ‡ Pathology and § Medicine, and || Gladstone Institute for Cardiovascular Disease, University of California, San Francisco, HSW-831, 505 Parnassus Avenue, San Francisco, California 94143-0524, USA

THROMBIN, a coagulation protease generated at sites of vascular injury, activates platelets, endothelial cells, leukocytes and mesenchymal cells^{1,2}. A G-protein-coupled receptor that is proteolytically activated by thrombin³ is a target for drug development aimed at blocking thrombosis, inflammation and proliferation. Here we show that although disruption of the thrombin-receptor (*tr*) gene in mice causes about half of the *tr*^{-/-} embryos to die at embryonic day 9-10, half survive to become grossly normal adult mice with no bleeding diathesis. Strikingly, *tr*^{-/-} platelets respond strongly to thrombin, whereas *tr*^{-/-} fibroblasts lose their ability to respond to thrombin. We conclude that the thrombin receptor plays an unexpected role in embryonic development, suggesting a possible new function for the 'coagulation' proteases themselves. Moreover, a second platelet thrombin receptor exists, and different thrombin receptors have tissue-specific roles. This may allow development of therapeutics that will selectively block thrombin's different cellular actions.

To define the roles of the cloned thrombin receptor *in vivo* and to answer definitively whether other receptors mediate some cellular responses to thrombin, we disrupted the cloned thrombin receptor (*tr*) gene in mice (Fig. 1). Surprisingly, *tr*^{-/-} progeny were born at about one-half of the expected rates. Litters from *tr*^{+/-} matings yielded 41 *tr*^{+/+}, 110 *tr*^{+/-}, and only 24 *tr*^{-/-} offspring ($P < 0.01$, χ^2 test). *tr*^{-/-} male X *tr*^{+/-} female matings yielded 98 *tr*^{+/+} and only 33 *tr*^{-/-} mice ($P < 0.01$, χ^2 test). At embryonic day 8.5 (E8.5), *tr*^{+/+} and *tr*^{+/-} embryos were morphologically indistinguishable and had beating hearts and a functional vitelline circulation (not shown). At E9.0, however, *tr*^{-/-} embryos were uniformly smaller than *tr*^{+/+} embryos (Fig. 1). By E9.5 this size

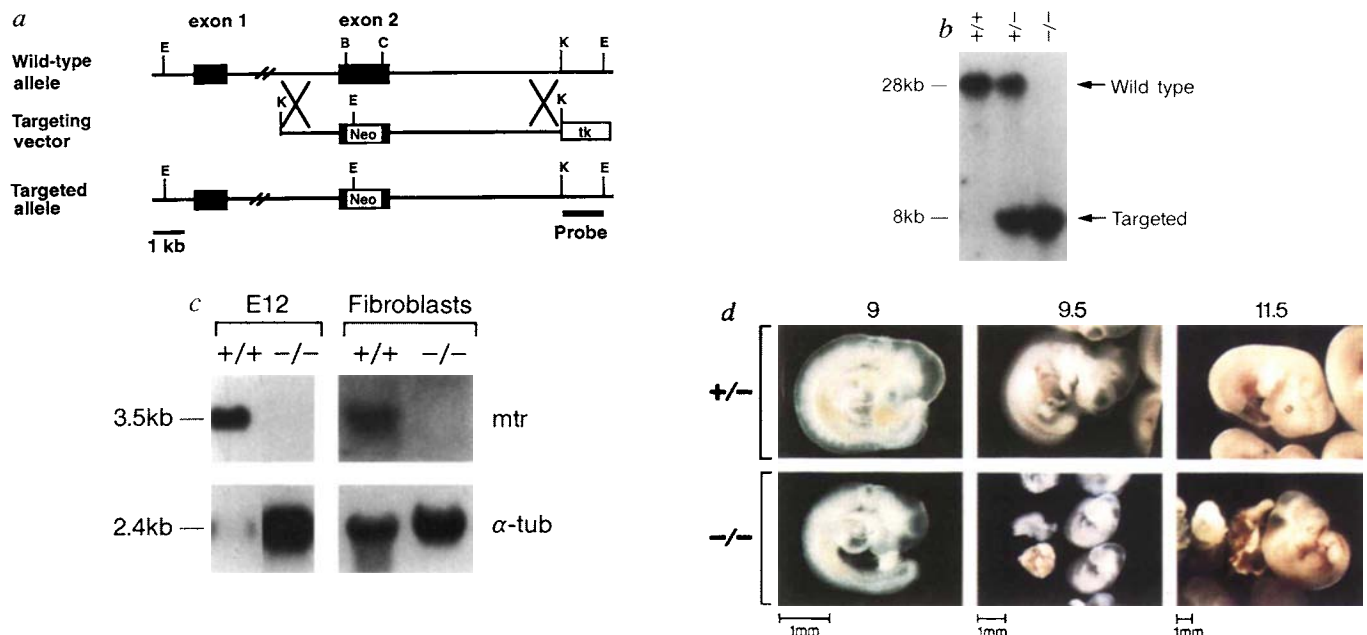


FIG. 1 Disruption of the thrombin-receptor gene in mice. *a*, Gene-targeting strategy: structure and partial restriction maps of *tr* gene, targeting construct and targeted allele, and location of the probe used for Southern analysis. Restriction endonuclease sites: E, *EcoRI*; B, *BspEI*; C, *Clal*; K, *KpnI*. *b*, Southern blot analysis of tail DNA from $tr^{+/+}$, $tr^{+/-}$, and $tr^{-/-}$ mice. The 28-kb wild-type and 8-kb targeted *EcoRI* fragments were identified using the 1.5-kb *KpnI*-*EcoRI* probe shown in *a*. *c*, Northern blot analysis of *tr* mRNA expression in $tr^{+/+}$ and $tr^{-/-}$ E12 embryos, and embryonic fibroblasts. *tr* mRNA was identified using the *BspEI*-*Clal* probe (upper panel); hybridization for α -tubulin was a control for lane loading (lower panel). *tr* mRNA was also absent from tissues of adult $tr^{-/-}$ mice (not shown). Sequencing of the product of reverse transcription and polymerase chain reaction (PCR) derived from $tr^{-/-}$ embryonic fibroblast RNA confirmed targeted disruption of the *tr* gene (not shown). *d*, Global growth retardation followed by death or recovery of $tr^{-/-}$ embryos. $tr^{+/+}$ or $tr^{+/-}$ embryos at E9.0, 9.5 or 11.5 are shown. Magnifications are the same for $tr^{-/-}$ and $tr^{+/+}$ embryos at each gestational age; 1 mm size standard is shown. No

differences between $tr^{+/+}$ and $tr^{+/-}$ embryos were detected (not shown). Results are representative of 4 litters examined at each time point.

METHODS. The mouse *tr* gene's exon 1 encodes the 5' untranslated sequence and signal peptide. Exon 2 encodes the entire mature receptor protein and 3' untranslated region¹². The targeting vector was constructed using a 7.5-kb *KpnI* genomic fragment and the PGK-Neo and PGK-tk cassettes from pPNT²⁵. The PGK-Neo cassette was inserted between *BspEI* and *Clal* sites in exon 2, thus replacing the gene segment encoding transmembrane domains 1–7. Culture and electroporation of RF8 ES cells (generated by R. Farese) and production of chimaeric mice was as described²⁶. Homologous recombination and germline transmission were confirmed by Southern blot and PCR. For PCR, primers to the inserted PGK-Neo cassette and to genomic sequence outside the 2.0-kb 5' flanking sequence in the targeting vector were used. $tr^{+/-}$ or $tr^{-/-}$ embryos were generated by mating $tr^{+/+}$ or $tr^{-/-}$ males to $tr^{-/-}$ females. Unless otherwise indicated, all results were obtained using mice that were 50% C57BL/6 and 50% 129/Sv strain.

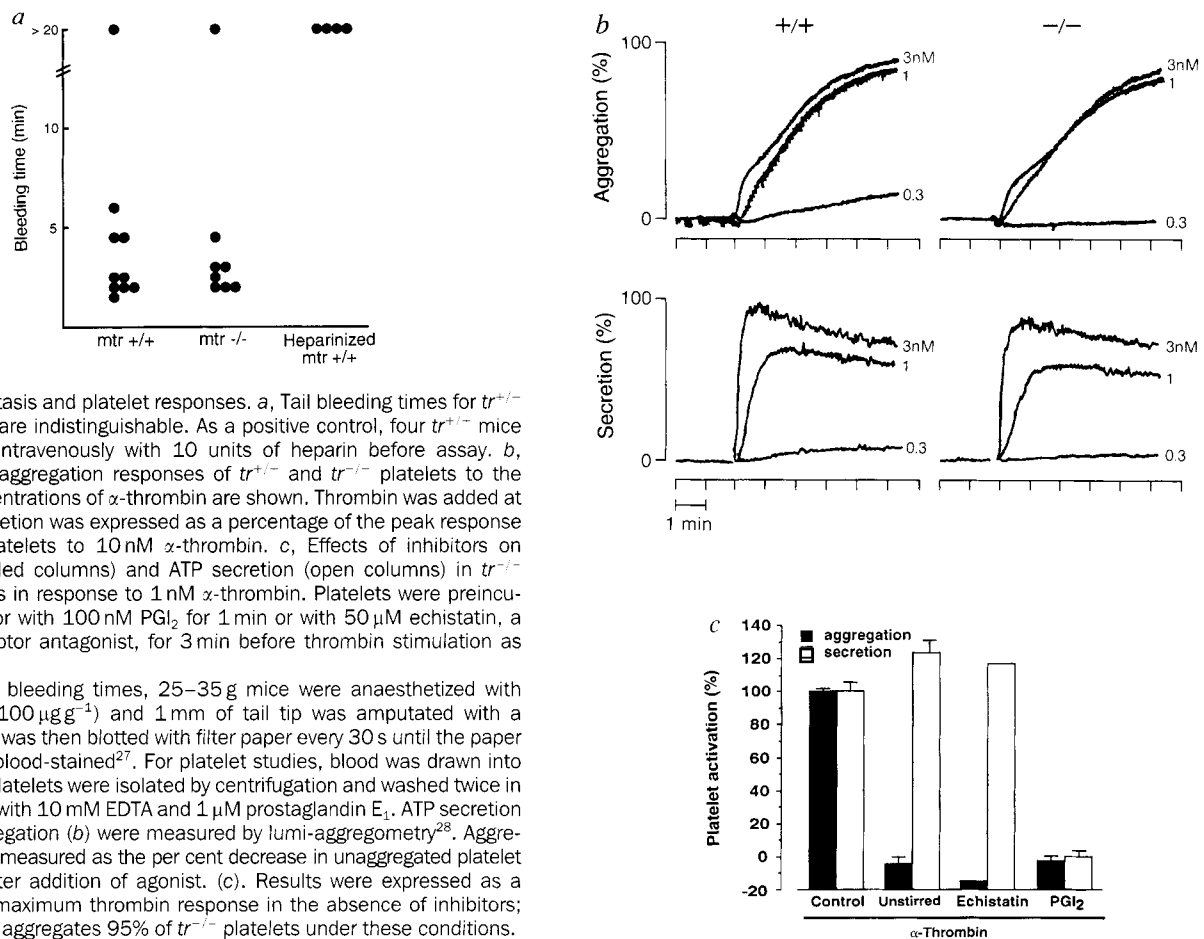
difference was marked, and about half of the $tr^{-/-}$ embryos had no heart beat (Fig. 1). A delay in placental development paralleled the delay seen in the $tr^{-/-}$ embryo proper at these time points. $tr^{-/-}$ embryos that survived the critical period from E9–10 had 'caught up' with their $tr^{+/+}$ counterparts by E11.5 (Fig. 1) and developed normally through to birth.

The cause of this embryonic death is unknown. Its partial nature is probably stochastic rather than genetic because it occurred even in an inbred 129/Sv background strain of mice (not shown). Failed haemostasis was not observed and would not be expected, given the platelet phenotype described below. Moreover, embryos lacking platelets⁴ or fibrinogen to support platelet aggregation⁵ survived to birth. At E9.0–9.5, *tr* messenger RNA was widely expressed in tissues critical for development at this time, including heart and blood vessels, haematopoietic cells, placenta, as well as brain and mesenchymal cells (ref. 6 and data not shown). Which of these tissues is responsible for the $tr^{-/-}$ embryonic lethality remains to be discovered. Regardless, our results reveal an unexpected developmental role for this receptor and raise new questions. Could thrombin produced upon invasion of maternal blood vessels be a signal for placental development, or might thrombin produced in the embryo have an unknown function? Alternatively, does another protease or even a peptide agonist activate the thrombin receptor during development?

Despite this partial embryonic loss, one-half of the $tr^{-/-}$ mice survived to birth, grew to maturity and were grossly normal. To our surprise, spontaneous bleeding was not seen in $tr^{-/-}$ mice and tail bleeding times, a platelet and thrombin-dependent *in vivo*

measure of haemostasis⁷, were indistinguishable from those of wild-type mice (Fig. 2*a*). Moreover, thrombin activated platelets from $tr^{-/-}$ or $tr^{+/+}$ mice equally (Fig. 2*b*). Strong secretion responses to thrombin were noted in platelets from $tr^{-/-}$ mice even in the absence of aggregation, so the secretion response was not aggregation-induced. Prostacyclin (PGI₂) blocked thrombin-induced secretion and aggregation in $tr^{-/-}$ platelets (Fig. 2*c*). Last, thrombin triggered rapid increases in cytoplasmic calcium in $tr^{-/-}$ platelets in the absence of extracellular calcium and aggregation (Fig. 3). These increases presumably reflect phosphoinositide hydrolysis and mobilization of calcium from intracellular stores. Thus thrombin can cause second messenger responses and functional activation in mouse platelets via a cell-surface receptor that is distinct from the cloned thrombin receptor.

What is the nature of this second receptor? The pharmacology of responses in $tr^{+/+}$ and $tr^{-/-}$ mouse platelets, human platelets, and *Xenopus* oocytes expressing cloned thrombin receptors was similar in many respects (Table 1). The protease inhibitor PPACK⁸ ablated thrombin's ability to activate $tr^{-/-}$ mouse platelets. γ -Thrombin, which is defective in its fibrinogen-binding exosite⁹, was two logs less potent than α -thrombin in activating $tr^{-/-}$ mouse platelets, and blockade of thrombin's fibrinogen-binding exosite with hirugen¹⁰ markedly right-shifted the dose response to thrombin. Thus thrombin's protease activity and its fibrinogen-binding exosite are important for activation of both the cloned and the second thrombin receptor. The cloned thrombin receptor and protease-activated receptor 2 (PAR2), which signals to trypsin but not thrombin, are clearly the result of gene



duplication^{11,12}. Whether the second thrombin receptor is another protease-activated receptor derived by gene duplication or something different remains to be determined.

What is the relative importance of the second thrombin receptor in human as opposed to mouse platelets? Synthetic peptides that mimic the new amino terminus that is unmasked when thrombin cleaves the cloned thrombin receptor are agonists for that receptor³. Such peptides fail to activate wild-type and $tr^{-/-}$ mouse platelets but readily activate human platelets (Table 1 and ref. 3). Similar species differences in agonist peptide potency have been reported but their basis is uncertain^{13,14}. The difference between mouse and human platelets was not accounted for by unresponsiveness of the cloned mouse thrombin receptor to agonist peptide; the mouse and human homologues responded similarly to agonist peptides or thrombin (Table 1). In addition, the failure of mouse platelets to respond to agonist peptides was not due to an inhibitor unique to mouse platelet preparations, because mixing of mouse and human platelets did not inhibit peptide-induced activation of human platelets (not shown). In the context of the observation that wild-type and $tr^{-/-}$ platelets respond identically to thrombin (Fig. 2), these results suggest that the cloned thrombin receptor plays little part in mouse platelet function. By contrast, the

cloned thrombin receptor does appear to be important for human platelets; thrombin-receptor-activating peptides cause a strong response in human platelets and thrombin-receptor antibodies attenuate their response to thrombin^{15–17}. However, thrombin is more effective than agonist peptide in eliciting certain responses in human platelet responses^{18–21}, so a second receptor may also operate in these cells.

In contrast to platelets, thrombin responses were lost in fibroblasts from $tr^{-/-}$ mice. Unlike their wild-type counterparts, fibroblasts derived from $tr^{-/-}$ embryos or adult lung failed to show

TABLE 1 Pharmacology of platelet activation versus cloned thrombin receptor activation

	Platelet ATP secretion		EC ₅₀	Cloned receptor activation	
	-/- Mouse	+/- Mouse		Mouse	Human
α -Thrombin (nM)	0.7	0.8	0.9	0.03	0.05
PPACK- α -thrombin (nM)	>1,000	>1,000	>1,000	>1,000	>1,000
γ -Thrombin (nM)	60	ND	60	2	5
α -Thrombin (nM) + hirugen	40	ND	30	25	40
SFLLRN (human) (μ M)	>200	>200	6	0.7	0.2
SFLLRNPS (mouse) (μ M)	>200	>200	180	3	3

Agonist-induced platelet ATP secretion was measured as in Fig. 1. Values of the half-maximal effective concentration (EC₅₀) were estimated from concentration–response curves using duplicate determinations at 4 different agonist concentrations. The fibrinogen-binding exosite inhibitor hirugen was incubated at 100 μ g ml⁻¹ with α -thrombin before addition to cells. To assess responses mediated by the cloned mouse and human receptors, cRNAs encoding these molecules were transcribed and injected into *Xenopus* oocytes. After 24 h in culture, oocytes were loaded with ⁴⁵Ca and agonist-induced release of ⁴⁵Ca was measured as an index of intracellular calcium flux²; EC₅₀ values were estimated as described²³. EC₅₀ values in the *Xenopus* system are uniformly lower than those obtained with platelets because of the high level of receptor expression achieved in the former²⁴. Note, however, that the relative potencies of the various agonists compared to α -thrombin is similar in the two systems.

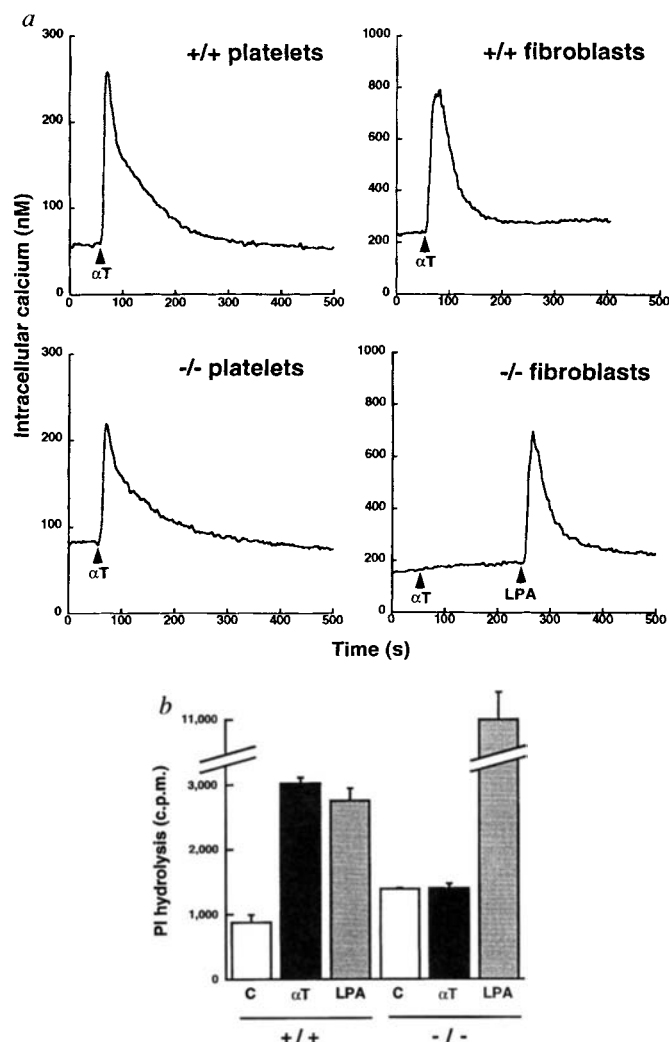


FIG. 3 Tissue-specific roles for the cloned thrombin receptor. *a*, Intracellular calcium measurements in $tr^{+/+}$ and $tr^{-/-}$ platelets and fibroblasts. Increases in cytosolic calcium in response to thrombin (α T; 10 nM) or lysophosphatidic acid (LPA; 1 μ M) were determined using Fura-2. Agonists were added at the indicated times (arrow). *b*, Agonist-induced phosphoinositide hydrolysis in $tr^{+/+}$ and $tr^{-/-}$ fibroblasts in response to thrombin (α T; 10 nM) or LPA (1 μ M).

METHODS. Platelets or fibroblasts were loaded with Fura-2AM dye at 37 °C for 30 min, suspended in calcium-free Tyrode's buffer containing 0.1 mM EGTA, and assayed using a Hitachi F2000 fluorometer²⁹. Absence of platelet aggregation during the calcium assay was confirmed by counting unaggregated platelets in a haemocytometer before and 4 min after addition of thrombin. Fibroblasts were cultured from wild-type or $tr^{-/-}$ E12.0 embryos or from the lungs of adult wild-type or $tr^{-/-}$ mice. Thrombin- or LPA-induced phosphoinositide hydrolysis was assessed as described³⁰, except that agonist stimulation lasted for 60 min. Data shown were obtained with lung fibroblasts at passages 20–24. Results were similar with fibroblasts from primary cultures and with passaged fibroblasts from two independent isolates. The magnitude of the LPA response differed for each fibroblast line tested.

thrombin-induced phosphoinositide hydrolysis and calcium mobilization (Fig. 3). Thrombin-induced mitogenic responses were also ablated in these cells (J. Trejo, A.J.C., and S.R.C., unpublished observations). The cloned thrombin receptor is thus critical for triggering classical thrombin-induced events in some mouse mesenchymal cells, a role that may contribute to the observed developmental phenotype.

In summary, the thrombin-receptor knockout mouse reveals an unexpected role for the cloned thrombin receptor in embryonic development and that the cloned receptor is critical in fibroblast signalling. The persistence of thrombin-induced responses in $tr^{-/-}$ platelets and its absence in $tr^{-/-}$ fibroblasts uncovers a second

platelet thrombin receptor and shows that different thrombin receptors have tissue-specific functions. Thrombin stimulation of mesenchymal cells may be important in normal wound healing and in pathological proliferative responses such as atherosclerosis and restenosis^{2,22}. With intact haemostatic function but absent fibroblast responses to thrombin, $tr^{-/-}$ mice provide an opportunity to test thrombin's role in *in vivo* models of these processes. Last, to the extent that multiple thrombin receptors probably also operate in humans, these studies suggest the possibility of developing pharmaceuticals that selectively block proliferative or haemostatic responses to thrombin. □

Received 20 March; accepted 18 April 1996.

- Fenton, J. W. *Sem. Hemost. Thromb.* **14**, 234–240 (1988).
- Coughlin, S. R. *Trends cardiovasc. Med.* **4**, 77–83 (1993).
- Vu, T.-K. H., Hung, D. T., Wheaton, V. I. & Coughlin, S. R. *Cell*, **64**, 1057–1068 (1991).
- Shivdasani, R. A. et al. *Cell*, **81**, 695–704 (1995).
- Suh, T. T. et al. *Genes Dev.* **9**, 2020–2033 (1995).
- Solfer, S. J., Peters, K. G., O'Keefe, J. & Coughlin, S. R. *Am. J. Path.* **144**, 60–69 (1993).
- Bevilacqua, L. et al. *Thromb. Res.* **71**, 3001–3015 (1993).
- Kettner, C. & Shaw, E. *Thromb. Res.* **14**, 969–973 (1979).
- Rydel, T. J. et al. *J. Biol. Chem.* **269**, 22000–22006 (1994).
- Skrzypczak, J. E. et al. *J. molec. Biol.* **221**, 1379–1393 (1991).
- Nystedt, S., Larsson, A. K., Aberg, H. & Sundelin, J. *J. Biol. Chem.* **270**, 5950–5955 (1995).
- Kahn, M. et al. *Molec. Med.* **2**, 349–357 (1996).
- Connolly, T. M. et al. *Thromb. Haemost.* **72**, 627–633 (1994).
- Derian, C. K., Santulli, R. J., Tomko, K. A., Haertlein, B. J. & Andrade-Gordon, P. *Thromb. Res.* **6**, 505–519 (1995).
- Hung, D. T., Vu, T.-K. H., Wheaton, V. I., Ishii, K. & Coughlin, S. R. *J. clin. Invest.* **89**, 1350–1353 (1992).
- Brass, L. F. et al. *J. Biol. Chem.* **267**, 13795–13798 (1992).
- Bahou, W. F. et al. *J. clin. Invest.* **91**, 1405–1413 (1993).
- Kinlough, R. R., Perry, D. W., Guccione, M. A., Rand, M. L. & Packham, M. A. *Thromb. Haemost.* **70**, 1019–1023 (1993).

- Goodwin, C. A. et al. *Biochem. biophys. Res. Commun.* **202**, 321–327 (1994).
- Lau, L. F., Pumiglia, K., Cote, Y. P. & Feinstein, M. B. *Biochem. J.* **391**, 391–400 (1994).
- Kramer, R. M. et al. *J. Biol. Chem.* **270**, 14816–14823 (1995).
- Fenton, J. Z., Ofose, F. A., Breznjak, D. V. & Hassouna, H. I. *Hemat. Oncol. Clin. North. Am.* **7**, 1107–1119 (1983).
- Gerszten, R. E. et al. *Nature* **368**, 648–651 (1994).
- Ishii, K. et al. *J. Biol. Chem.* **269**, 1125–1130 (1994).
- Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153–1163 (1991).
- Hogan, B., Beddington, R., Constantini, F. & Lacy, E. *Manipulating the Mouse Embryo* 1–196–290 (Cold Spring Harbor Laboratory Press, New York, 1994).
- Carmeliet, P. et al. *J. clin. Invest.* **92**, 2756–2760 (1993).
- Charo, I. F., Feinman, R. D. & Detwiler, T. C. *J. clin. invest.* **60**, 866–873 (1977).
- Lages, B. & Weiss, H. *J. Blood* **83**, 2549–2559 (1994).
- Hung, D. T., Wong, Y. H., Vu, T.-K. H. & Coughlin, S. R. *J. Biol. Chem.* **353**, 20831–20834 (1992).

ACKNOWLEDGEMENTS. We thank E. Sande for technical assistance, J. Fenton of Albany Medical College for α -thrombin and its derivatives, and H. Bourne and I. Charo of UCSF for critical reading of this manuscript.

CORRESPONDENCE and requests for materials should be addressed to S.R.C. (e-mail: shaun_coughlin.src@quickmail.usc.edu).

Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot

Centre de Recherche Cerveau & Cognition, UMR 5549, 31062 Toulouse, France

How long does it take for the human visual system to process a complex natural image? Subjectively, recognition of familiar objects and scenes appears to be virtually instantaneous, but measuring this processing time experimentally has proved difficult. Behavioural measures such as reaction times can be used¹, but these include not only visual processing but also the time required for response execution. However, event-related potentials (ERPs) can sometimes reveal signs of neural processing well before the motor output². Here we use a go/no-go categorization task in which subjects have to decide whether a previously unseen photograph, flashed on for just 20 ms, contains an animal. ERP analysis revealed a frontal negativity specific to no-go trials that develops roughly 150 ms after stimulus onset. We conclude that the visual processing needed to perform this highly demanding task can be achieved in under 150 ms.

Neurophysiological measurements of the latencies of selective visual responses can be used to provide estimates of visual processing time³. For example, it is known that higher-order visual areas such as the primate superior temporal sulcus contain neurons that can respond selectively to faces with latencies of ~100 ms⁴⁻⁶. In humans, face-selective evoked potentials have been demonstrated using both surface ERP recordings^{7,8} and implanted intracerebral electrodes⁹⁻¹¹. Such potentials typically peak at ~200 ms after stimulus onset, but may start as early as 140 ms. It is unclear, however, whether such latencies are typical of visual processing in general. One problem is that face processing may involve highly specialized and optimized neural pathways, and although there have been a few reports of early differential responses to other stimuli, including words^{10,12,13} and line drawings¹³, no previous ERP studies have attempted to measure processing times for more natural scenes. A second problem is that the existence of a short latency differential response does not imply that visual processing has been completed—responses to faces, for example, could correspond to an early processing stage such as ‘structural encoding’. One can tackle this problem by using a task that requires the subject to make some sort of categorical judgment about the stimulus, an approach that has been used in a variety of studies using not only photographs of faces¹⁴⁻¹⁷, but also both line drawings^{18,19} and photographs²⁰ of everyday objects. However, differential effects in such tasks occur at considerably longer latencies, typically involving the N400 component of the ERP.

The present study used a task that provides a very serious challenge to the processing capacities of the human visual system. Subjects performed a go/no-go categorization in which they had to decide on the basis of a 20-ms presentation whether an image contained an animal or not. Earlier studies using rapid sequential visual presentation (RSVP) had shown that subjects can detect photographs of animals in a string of images at high presentation rates²¹, but this is the first time such a scene categorization task has been performed using ERPs. We used a set of over 4,000 commercially available colour photographs, of which roughly half were used as targets and included a wide range of animals in their natural environments (mammals, birds, reptiles, fish); the remainder were distractors that included pictures of forests, mountains and lakes, as well as buildings, flowers and fruit. As in a number of other studies^{8,9,11,16,19}, each stimulus was only ever seen once, thus eliminating the possibility of stimulus-specific learning effects.

Despite the very high demands made on the visual system by

such a task (the subjects had no *a priori* information about the type of animal to look for, its position or size, or even the number of animals present), performance was remarkably good. The average proportion of correct responses was 94%, with one of the fifteen subjects achieving 98% correct responses. The median reaction times on ‘go’ trials was 445 ms, although this value varied considerably between subjects, from a minimum of 382 ms to as much as 567 ms (Fig. 1). This remarkable level of performance was possible despite the very brief presentations, which effectively rule out the use of eye movements during image processing.

Whereas the behavioural reaction times put an upper limit on the time required for visual processing, the analysis of event-related potentials provided a much stronger constraint. By comparing average brain potentials generated on correct ‘go’ trials with those generated on correct ‘no-go’ trials, we were able to demonstrate that the two potentials diverge very sharply at ~150 ms after stimulus onset. The effect was particularly clear at frontal recording sites, and was characterized by a nearly linear increase in the voltage difference over the following 50 ms or so, the potential being more negative on no-go trials (Fig. 2). All 15 subjects showed the effect (Fig. 3), and although the onset latency varied somewhat between subjects, the differences were very minor compared with the very large differences in behavioural reaction times. Furthermore, there was no correlation whatsoever between behavioural reaction time and the onset latency for the differential response. This makes it unlikely that the differential

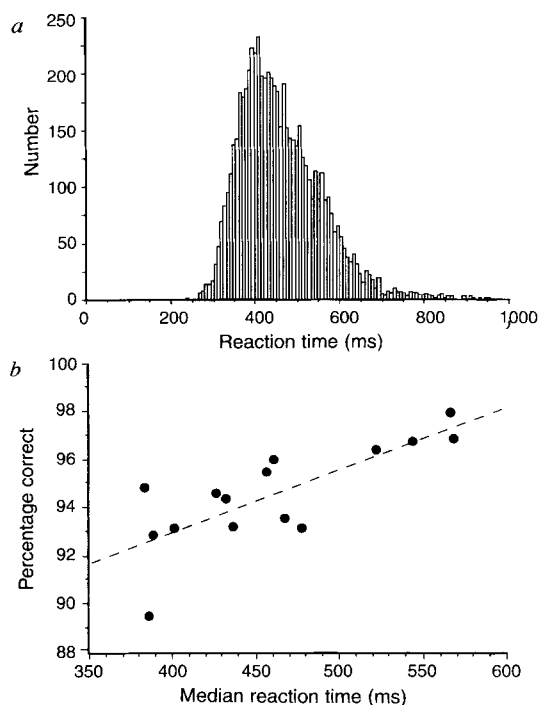
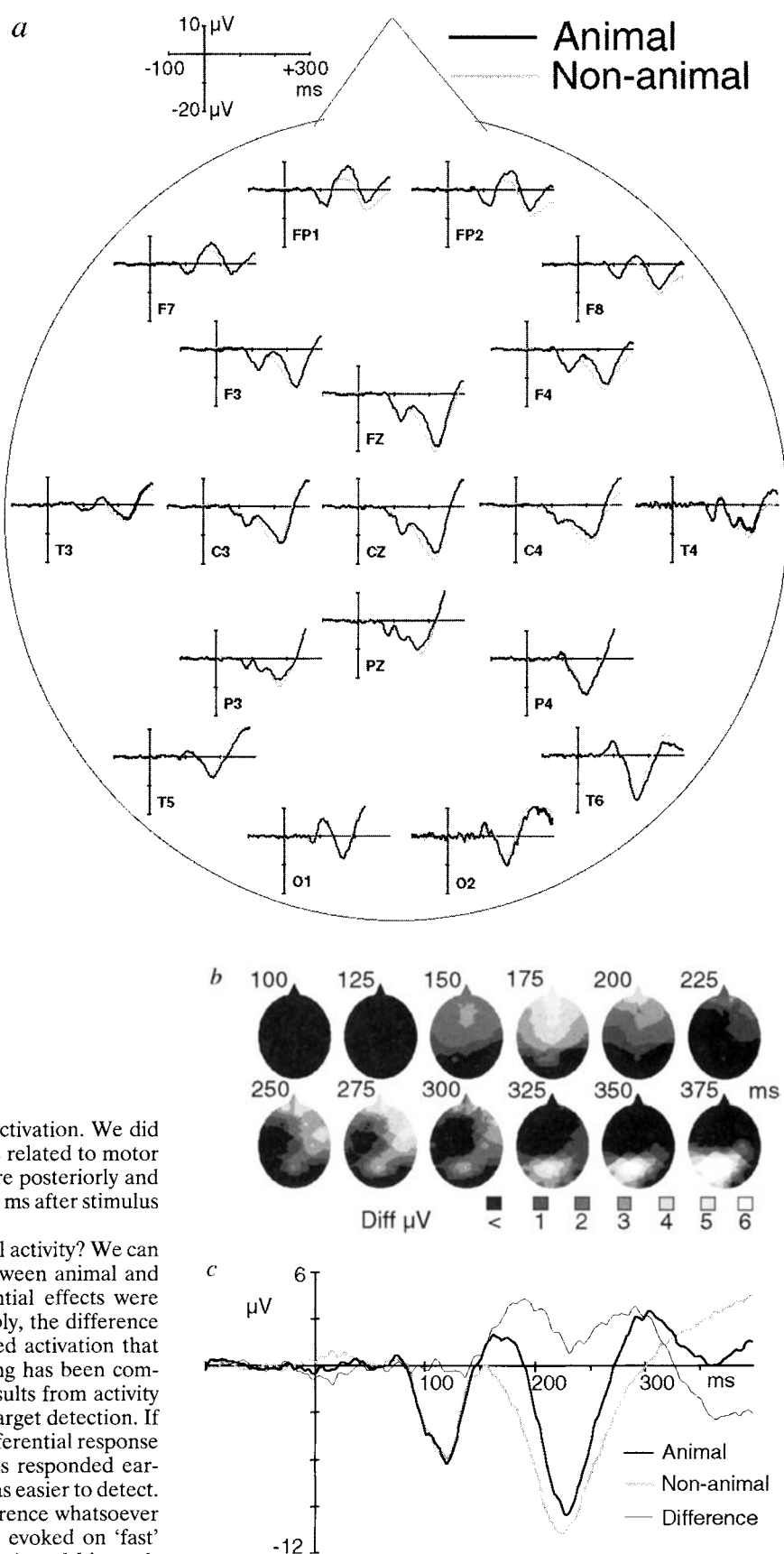


FIG. 1 Behavioural performance measures. *a*, Distribution of reaction times on target trials for the 15 subjects participating in the study (7 males and 8 females, aged between 22 and 45 years of age), each of whom performed at least 700 trials (range, 700–2,000). Subjects pressed a button to start a sequence of trials and, after an interval of 1 to 2 seconds, an image was presented at the centre of the screen. A small fixation cross was present before and after stimulus presentation. Subjects were instructed to release the button if they saw an animal (‘go’ trials), and to keep their finger on the button otherwise (‘no-go’ trials). Target and distractor trials were presented at random, with roughly equal probability in blocks of 100 trials. All the images were 384 by 256 pixels in size and were presented for 20 ms (two frames at 100 Hz) using a Cambridge Vision Research VSG 2/2 graphics board mounted in a PC compatible computer. *b*, Accuracy as a function of median reaction time for each of the 15 subjects. The dashed line plots a linear regression ($r = 0.623$) and indicates the presence of a speed–accuracy trade-off.

FIG. 2 Evoked potential data for one subject. *a*, Event-related potentials plotted for a 400-ms period starting 100 ms before stimulus onset. Solid lines plot the average response on correct target trials; grey lines plot averages for correct distractor trials. Data were obtained from one subject during 4 separate sessions (a total of 1,580 trials). The two averages overlap until ~ 150 ms, at which point a clear difference emerges between the potentials on animal and non-animal trials which was clearest at frontal recording sites. Recordings were made using a 20-electrode Electrocap bonnet in the 10–20 configuration connected to a Neuroscan SynAmps system sampling at 1,000 Hz. Signals were digitally filtered using a low pass filter with a cut-off frequency of 100 Hz and a notch filter to remove 50 Hz mains interference. Potentials on individual trials were baseline corrected on the basis of the 100 ms preceding stimulus onset, and any trials contaminated by eye movement artefacts were excluded from the analysis. Data analysis was performed using the SCAN software suite. *b*, Two-dimensional plots of the difference between correct 'go' and 'no-go' trials averaged over 25-ms time slices, starting at 100 ms. No difference can be seen in the first two slices, but from 150 ms onwards a clear difference appears which affects all the frontal electrodes. At around 325 ms, there are clear signs of a lateralized difference at more posterior sites which probably reflects motor activity in this right-handed subject. *c*, Event-related potentials on target (thick black lines) and distractor (grey) trials calculated by averaging the responses for all seven frontal electrodes (FP1, FP2, F3, F4, F7, F8 and FZ) for this one subject. The difference curve (thin black line) demonstrates the sharp onset of the differential response which is more negative on no-go trials. Using the statistical procedure proposed by Rugg *et al.* for determining the onset of the differential effect²⁸ (at least 15 consecutive *t*-test values exceeding the 0.05 level of significance) the first significant effect was seen for electrode F8 at 152 ms (*d.f.* = 1,578). All seven electrodes reached significance by 157 ms ($2.20 < t < 3.74$), and the level of significance increased monotonically to reach a peak at 186 ms (mean *t*-score for the seven electrodes, 6.72).

effect seen at frontal sites is related to motor activation. We did see lateralized differential activity that could be related to motor preparation (Fig. 2*b*) but this was localized more posteriorly and did not occur until considerably later (300 to 400 ms after stimulus onset).

What could be producing the early differential activity? We can rule out some simple systematic difference between animal and non-animal images because the same differential effects were seen for a very wide range of images. Presumably, the difference must be related to some sort of decision-related activation that occurs only once the necessary visual processing has been completed. One possibility is that the difference results from activity that is specific to 'go' trials, perhaps related to target detection. If target detection was indeed critical, then the differential response should start earlier on trials where the subjects responded earlier—that is, on trials where the target animal was easier to detect. However, Fig. 3*c* shows that there was no difference whatsoever between the latency of the differential activity evoked on 'fast' trials and 'slow' trials. The most plausible explanation of this result is that the difference is not generated by 'go'-related neural activity, but rather by neural activity that is specifically generated on 'no-go' trials. This could reflect a role for frontal areas in inhibiting inappropriate behavioural responses, an interpretation supported by a number of earlier studies that reported frontal



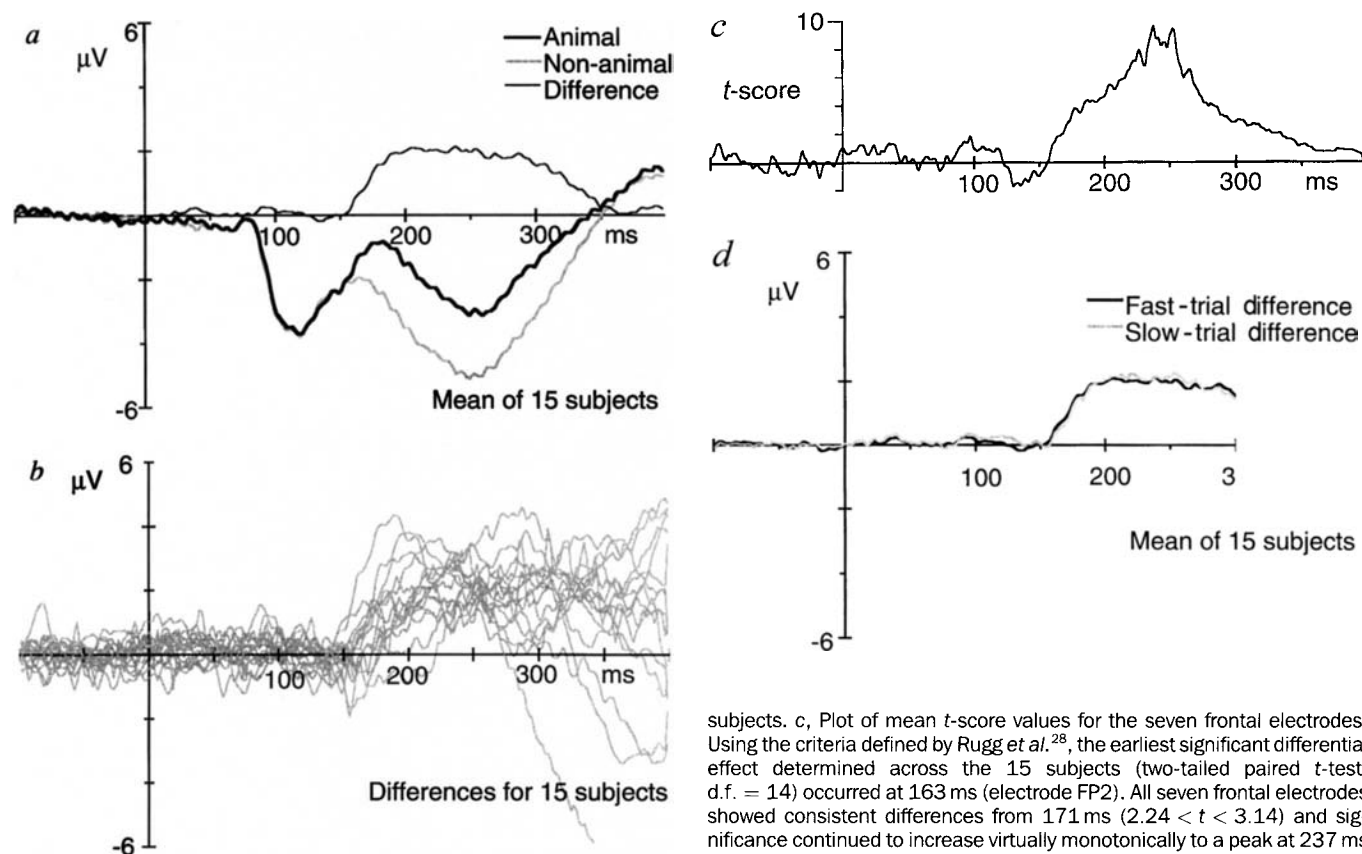


FIG. 3 Event-related potentials for 15 subjects. *a*, As Fig. 2c, but averaged over all 15 subjects. *b*, Average difference curves for the seven frontal electrodes plotted separately for each of the 15 subjects. Note that all subjects show a similar difference function, more negative on 'no-go' trials, and that the onset of the differential response is relatively constant across

subjects. *c*, Plot of mean *t*-score values for the seven frontal electrodes. Using the criteria defined by Rugg *et al.*²⁸, the earliest significant differential effect determined across the 15 subjects (two-tailed paired *t*-test, d.f. = 14) occurred at 163 ms (electrode FP2). All seven frontal electrodes showed consistent differences from 171 ms ($2.24 < t < 3.14$) and significance continued to increase virtually monotonically to a peak at 237 ms where the mean *t*-score was 9.58. *d*, Effect of target difficulty on the differential response. Two separate difference functions are shown. The first ('fast-trial difference') was calculated using those go trials where the subjects reaction time was faster than the median, whereas the second ('slow-trial difference') used those trials where the subject was slower. The fact that the two curves overlap virtually perfectly indicates that the differential response is probably the result of activity specific to 'no-go' trials.

activity specific to 'no-go' trials at around the same latency, but where the visual processing requirements were much less demanding than in this study^{22,23}. It is clear, however, that additional work using techniques such as fMRI and/or source analysis will be needed to determine the precise structures involved in generating the differential response.

The presence of 'no-go' specific activity at frontal recording sites at 150 ms implies that a great deal of visual processing must have been completed before this time. Indeed, although activity related to target detection could be explained relatively easily (the presence of an eye or feathers would be enough to decide that an

animal is present), 'no-go' specific activity implies that the visual system has already performed enough processing to conclude that no animal is present anywhere in the image. It therefore seems clear that the very rapid processing seen previously in the case of faces also occurs in the case of much more complex scene analysis. Quite how the human visual system achieves such a phenomenal amount of computation in such a short time is clearly a challenge for current theories of object vision²⁴⁻²⁶, but given the large number of processing stages involved in primate visual system, it seems likely that much of this processing must be based on essentially feed-forward mechanisms^{3,4,27}. □

Received 21 November 1995; accepted 22 March 1996.

- Potter, M. C. & Faulconer, B. A. *Nature* **253**, 437-438 (1975).
- Rugg, M. D. & Coles, M. G. H. *Electrophysiology of Mind: Event-related Brain Potentials and Cognition* (Oxford University Press, Oxford, 1995).
- Thorpe, S. J. & Imbert, M. in *Connectionism in Perspective* (eds Pfeifer, R., Schreier, Z., Fogelman-Soulie, F. & Steels, L.) 63-92 (Elsevier, Amsterdam, 1989).
- Oram, M. W. & Perrett, D. I. *J. Neurophysiol.* **68**, 70-84 (1992).
- Perrett, D. I., Rolls, E. T. & Caan, W. *Exp. Brain Res.* **47**, 329-342 (1982).
- Rolls, E. T. & Tovee, M. J. *Proc. R. Soc. Lond. B* **257**, 9-15 (1994).
- Jeffreys, D. A. & Tukmach, E. S. *Exp. Brain Res.* **91**, 340-350 (1992).
- Seeck, M. & Grüsser, O. J. *Exp. Brain Res.* **92**, 338-349 (1992).
- Halgren, E., Baudena, P., Heit, G., Clarke, M., Marinkovic, K. & Clark, M. J. *Physiol., Paris* **88**, 1-50 (1994).
- Allison, T., McCarthy, G., Nobre, A., Puce, A. & Belger, A. *Cereb. Cortex* **4**, 544-554 (1994).
- Allison, T. *et al.* *J. Neurophysiol.* **71**, 821-825 (1994).
- Nobre, A. C., Allison, T. & McCarthy, G. *Nature* **372**, 260-263 (1994).
- Rudell, A. P. *EEG Clin. Neurophysiol.* **83**, 77-82 (1992).
- Barrett, S. E., Rugg, M. D. & Perrett, D. I. *Neuropsychologia* **26**, 105-117 (1988).

- Barrett, S. E. & Rugg, M. D. *Neuropsychologia* **27**, 913-922 (1989).
- Holcomb, P. J. & MacPherson, W. B. *Brain Cogn.* **24**, 259-276 (1994).
- Bobes, M. A., Valdes Sosa, M. & Olivares, E. *Brain Cogn.* **26**, 1-22 (1994).
- Barrett, S. E. & Rugg, M. D. *Brain Lang.* **38**, 424-437 (1990).
- Barrett, S. E. & Rugg, M. D. *Brain Cogn.* **14**, 201-212 (1990).
- Pratarelli, M. E. *Brain Cogn.* **24**, 137-157 (1994).
- Intraub, H. J. *Exp. Psychol. Hum. Perc. Perf.* **7**, 604-610 (1981).
- Gemba, H. & Sasaki, K. *Neurosci. Lett.* **101**, 263-268 (1989).
- Sasaki, K., Gemba, H., Nambu, A. & Matsuzaki, R. *Neurosci. Res.* **18**, 249-252 (1993).
- Vetter, T., Hurlbert, A. & Poggio, T. *Cereb. Cortex* **5**, 261-269 (1995).
- Bülthoff, H. H., Edelman, S. Y. & Tarr, M. *Cereb. Cortex* **5**, 247-260 (1995).
- Hummel, J. E. & Biederman, I. *Psychol. Rev.* **99**, 480-517 (1992).
- Celebrini, S., Thorpe, S., Trotter, Y. & Imbert, M. *Vis. Neurosci.* **10**, 811-825 (1993).
- Rugg, M. D., Doyle, M. C. & Wells, T. J. *cog. Neurosci.* **7**, 209-227 (1995).

ACKNOWLEDGEMENTS. This work was supported by the Human Capital and Mobility program, the CNRS and the Conseil Régional Midi-Pyrénées.

CORRESPONDENCE and requests for materials should be addressed to S.T. (e-mail: thorpe@cerco.ups-tlse.fr).

Target-cell-specific concentration of a metabotropic glutamate receptor in the presynaptic active zone

Ryuichi Shigemoto^{*†}, Akos Kulik^{*}, J. David B. Roberts^{*}, Hitoshi Ohishi[†], Zoltan Nusser^{*}, Takeshi Kaneko[†] & Peter Somogyi^{*}

^{*} Medical Research Council, Anatomical Neuropharmacology Unit, Department of Pharmacology, Oxford University, Mansfield Road, Oxford OX1 3TH, UK

[†] Department of Morphological Brain Science, Faculty of Medicine, Kyoto University, Sakyo-ku, Kyoto 606, Japan

THE probability of synaptic neurotransmitter release from nerve terminals is regulated by presynaptic receptors responding to transmitters released from the same nerve terminal or from terminals of other neurons. The release of glutamate, the major excitatory neurotransmitter, is suppressed by presynaptic autoreceptors¹⁻³. Here we show that a metabotropic glutamate receptor (mGluR7) in the rat hippocampus is restricted to the presynaptic grid, the site of synaptic vesicle fusion. Pyramidal cell terminals presynaptic to mGluR1 α -expressing interneurons have at least a ten-fold higher level of presynaptic mGluR7 than terminals making synapses with pyramidal cells and other types of interneuron. Distinct levels of mGluR7 are found at different synapses made by individual pyramidal axons or even single boutons. These results raise the possibility that presynaptic neurons could regulate the probability of transmitter release at individual synapses according to the postsynaptic target.

The release of glutamate is suppressed by L-2-amino-4-phosphonobutyrate (AP4)-sensitive, presynaptic metabotropic glutamate receptors⁴⁻⁶, which have been postulated to serve as autoreceptors^{1,3}. We used an antibody specific to mGluR7 (Fig. 1c), the only known AP4-sensitive mGluR expressed abundantly in the hippocampus⁷⁻⁹, to establish its location by means of a high-resolution immunodetection method¹⁰. Light microscopy revealed extensive distribution of mGluR7 immunoreactivity throughout the hippocampal dendritic fields¹¹, some neuronal profiles having strongly immunoreactive puncta, prominent over the weaker neuropile labelling (Fig. 1a). Double immunolabelling with an antibody specific to mGluR1 α (Fig. 1c) showed that these profiles corresponded to an interneuron population¹⁰ selectively expressing mGluR1 α in stratum oriens of the CA1 area (Fig. 1b), the CA3 area and the hilus (not shown). Electron-microscopic analysis demonstrated that immunoparticle labelling for mGluR7 was highly restricted to the presynaptic membrane specialization of axon terminals making type I (asymmetrical) synapses (Fig. 2a), which are known to contain glutamate. All terminals that make type I synapses with dendrites that express mGluR1 α had a much higher density of the mGluR7 labelling (Fig. 2a, b) than terminals making synapses with mGluR1 α -immunonegative dendritic shafts (Fig. 2b) or pyramidal-cell spines (not shown). Measurement of the immunoparticle density at individual synapses (Fig. 2c) indicated a highly significant difference between synapses on mGluR1 α -positive ($76 \pm 16\%$, mean $\pm$ s.d., $n = 16$) and negative ($7.1 \pm 7.2\%$, $n = 44$) neuronal elements (Mann-Whitney test, $Z = -5.9$, $P < 0.0001$). It is unlikely that the

differential immunolabelling is caused by a selective lack of access of antibodies to the presynaptic grid in synapses on spines and some dendritic shafts because the differential pattern was observed in the surface of the tissue and in detergent-treated material for fluorescence microscopy. The expression of a selectively high level of mGluR7 does not depend on the presence of mGluR1 α *per se*, as the pattern was the same in genetically altered mice lacking mGluR1 α expression (R. Shigemoto, unpublished observation).

The presence of nerve-terminal populations with two distinct levels of presynaptic mGluR7 raises the question of whether the terminals originate from two different populations of presynaptic neuron, or whether both types of terminal occur along a single axon with the difference depending on the identity of the postsynaptic neuron. This was investigated by identifying individual axons of pyramidal cells using *Phaseolus vulgaris* leucoagglutinin (PHAL) injected in the CA3 (Fig. 3) and CA1 areas and double labelling the material for mGluR7 with immunoparticles. Single, PHAL-labelled associational axons of CA3 pyramidal cells occasionally contacted mGluR7-decorated dendritic shafts (Fig. 3a), which originate from mGluR1 α -positive interneurons (Figs 1 and 2). Electron-microscopic analysis showed that in 9 of 18 tested cases these PHAL-labelled boutons formed synapses with the dendrites, and the presynaptic grid was densely labelled with immunoparticles for mGluR7 (Fig. 3c). In 7 cases in the CA3

FIG. 1 Correlated distribution of strong immunoreactivity for metabotropic glutamate receptors mGluR7 and mGluR1 α in rat hippocampus. Punctate immunofluorescence for mGluR7 (a) decorates mGluR1 α -immunopositive soma and dendrites (b) of interneurons in stratum oriens/aleus of the CA1 area. Double immunolabelling for mGluR7 and mGluR1 α was photographed with different filters for Texas Red (mGluR7) or fluorescein (mGluR1 α). Scale bar, 20 μ m. c, Immunoblots of a membrane fraction from the whole brain with antibodies to mGluR1 α (lane 1) and mGluR7 (lane 3), showing major immunoreactive bands of relative molecular mass 145,000 (M_r 145K) and 102,000 (M102K), respectively. For the immunoblots in lane 2 and 4, antibodies were preabsorbed with the corresponding fusion proteins before immunoreaction. M_r markers are indicated on the left.

METHODS. Guinea-pig antibodies to mGluR1 α and rabbit antibodies to mGluR7 were raised using *trpE*-mGluR fusion proteins³⁰. cDNA fragments encoding C-terminal amino-acid residues of rat mGluR1 α ¹ (859–1,199) and mGluR7 (ref. 8) (874–915) were inserted into pATH3 vectors and the *trpE*-mGluR proteins were overexpressed, purified and used for immunization. After removing antibodies to the *trpE* protein, antibodies to mGluR sequences were purified with antigen columns. For double-immunofluorescence histochemistry, hippocampal sections 20 μ m thick from rat, fixed by perfusion with 4% paraformaldehyde¹³, were reacted with antibodies to mGluRs (0.5–1.0 μ g ml⁻¹), then with fluorescein isothiocyanate-conjugated anti-guinea-pig-IgG antibody and biotinylated anti-rabbit-IgG antibody combined with Texas-Red-conjugated avidin. When either of the primary antibodies was omitted, no fluorescence signal of the omitted primary antibody was observed (not shown). For immunoblots, a crude membrane fraction prepared from the rat brain was separated on 7% SDS-PAGE and transferred to polyvinylidene difluoride membrane¹³. The membranes were reacted with antibodies to mGluRs (1.0 μ g ml⁻¹), and the bands were visualized with alkaline phosphatase-labelled secondary antibodies.

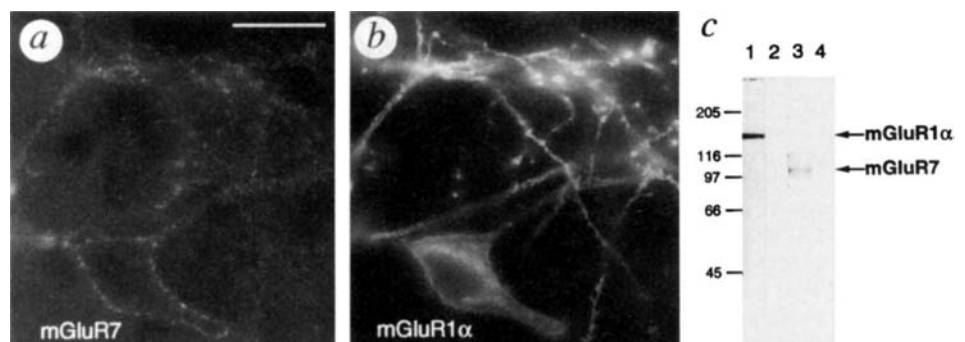
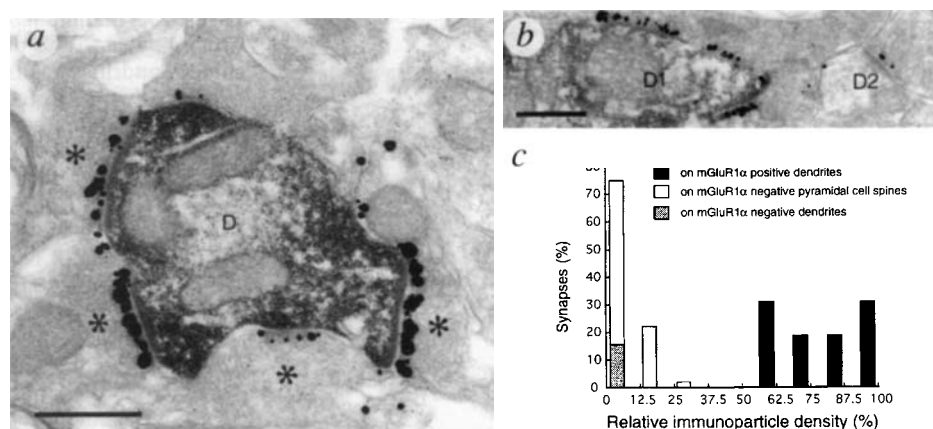


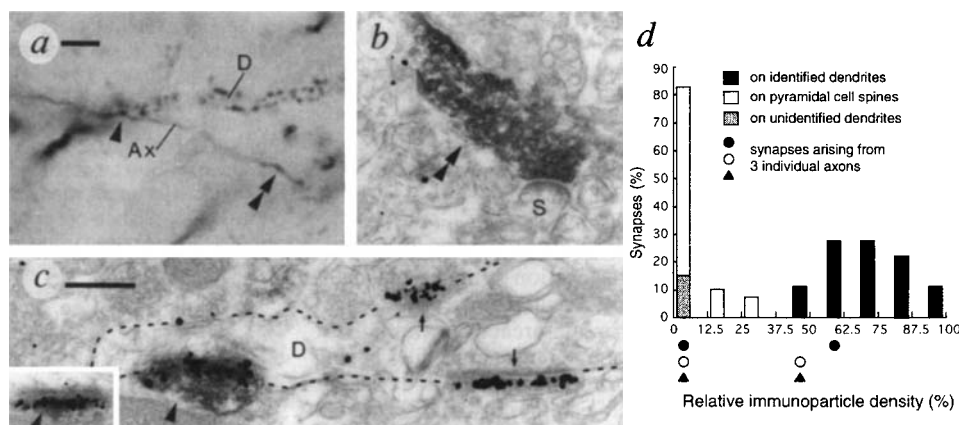
FIG. 2 Expression of mGluR7 in the presynaptic grid of terminals on mGluR1 α immunopositive neurons. *a, b*, Electron micrographs of terminals labelled for mGluR7 (particles) and dendrites (D, D1) labelled for mGluR1 α (peroxidase product) in the CA3 area. The transmitter release site of terminals (asterisks) making synaptic junctions with mGluR1 α -immunoreactive dendrites (D, D1) are heavily labelled for mGluR7 with silver-enhanced immunogold particles. Terminals forming synapses on a mGluR1 α -negative dendrite (D2) or spines (not shown) are labelled only weakly, if at all. Scale bars, 0.4 μ m. *c*, Measurement of presynaptic mGluR7 immunoreactivity in the CA3 area showed that asymmetrical synapses on mGluR1 α -positive dendrites (black column) have much higher density of receptor ($P < 0.0001$) than those on dendrites of other interneurons (grey column) and on pyramidal cell spines (white column).

METHODS. Wistar rats were anaesthetized (sodium pentobarbital, 150 mg per kg, intraperitoneal) and perfused with a fixative containing 4% paraformaldehyde, 0.05% glutaraldehyde, and 0.2% picric acid in 0.1 M phosphate buffer. Vibratome sections (50 μ m thick) were incubated with antibodies (0.5–1.0 μ g ml $^{-1}$) to mGluR1 α (guinea-pig) and mGluR7 (rabbit), then with biotinylated anti-guinea-pig-IgG antibody and 1.4 nm gold-coupled anti-rabbit-IgG antibody. After silver enhancement, sections were reacted with the avidin-biotinylated peroxidase complex (Vector), and mGluR1 α immunoreactivity was visualized with diaminobenzidine tetrahydrochloride. When either of the primary antibodies was omitted, no signal



for the omitted antibody was observed (not shown). Measurements of immunoparticle density were taken from electron micrographs having mGluR1 α -positive dendrites in the centre and covering an area of 26 μ m 2 . Every synaptic profile that had a well-defined synaptic cleft and thick postsynaptic density was measured. Particle density was determined at presynaptic grids by dividing the total area of particles (measured with OptiLab image-processing software) by the length of the synaptic specialization. Density values from two consecutive sections were averaged and normalized by taking the highest value within each sample as 100%. The numbers of synapses are normalized within categories according to the presence or absence of postsynaptic mGluR1 α immunoreactivity.

FIG. 3 Individual presynaptic terminals along the same pyramidal cell axon have different levels of presynaptic mGluR7, depending on the identity of the postsynaptic neuron. Corresponding light (*a*) and electron (*b, c*) micrographs showing a pyramidal cell axon collateral (Ax) and mGluR7 immunogold labelling along dendrites (one is labelled D) in the CA3 area. The axon is likely to originate from a pyramidal cell, as identified by anterograde labelling with PHAL (visualized with immunoperoxidase) injected in the ipsilateral CA3 stratum radiatum. One PHAL-labelled bouton (single arrowhead) makes a synapse that is heavily labelled by immunoparticles for mGluR7 (*c*; and inset from a tilted consecutive section), whereas the other PHAL-labelled bouton (double arrowheads) makes an immunonegative synapse (*b*) with a dendritic spine (S) of a pyramidal cell. Other synapses (arrows in *c*) on the same dendrite (D, broken lines) are also heavily labelled for mGluR7. Scale bars: *a*, 3 μ m; *b, c*, same magnification, 0.3 μ m. *d*, Distinct levels of presynaptic mGluR7 immunoreactivity in individual boutons along 3 identified pyramidal cell axons (symbols) according to the identity of postsynaptic targets. The level of mGluR7 in PHAL-labelled terminals was compared to that of other asymmetrical synapses around the identified boutons in the CA3 area. Asymmetrical synapses on dendrites (black columns), which received mostly heavily labelled terminals,



have a much higher density of receptor ($P < 0.0001$) than those on dendrites of unidentified interneurons (grey column) and spines of pyramidal cells (white columns). Different boutons of the same axon contribute to both populations of synapses.

METHODS. Double labelling for mGluR7 (immunogold) and PHAL (immunoperoxidase) was performed as described in Fig. 2, but biotinylated goat anti-PHAL antibody (Vector) was used instead of the mGluR1 α and secondary antibodies. Relative immunoparticle densities were measured from three blocks of two animals. For quantification see Fig. 2.

area the same axons could be followed in serial sections to a consecutive synaptic bouton (Fig. 3, double arrowheads), making a synapse with a pyramidal-cell dendritic spine. The presynaptic grid of the latter boutons was only weakly labelled, or not labelled at all, for mGluR7 (Fig. 3*b*). Measurements of three pairs of synapses (Fig. 3*d*), together with all other synapses found around the PHAL-labelled boutons, demonstrate that a single axon contributes individual boutons to two populations of synapses (Mann–Whitney test, $Z = -6.1$, $P < 0.0001$), having either high ($69 \pm 17\%$, mean $\pm$ s.d., $n = 18$) or low ($7.3 \pm 10.0\%$, $n = 41$) levels of presynaptic mGluR7. Similar results were obtained in

the CA1 area, where three different axons of CA1 pyramidal cells (2–4 boutons measured from each), identified by PHAL labelling, were found to contribute to populations of synapses with either high ($54 \pm 21\%$, $n = 30$) or low ($7.1 \pm 8.5\%$, $n = 59$) mGluR7 density (Mann–Whitney test, $Z = -7.5$, $P < 0.0001$). Individual synapses with high or low mGluR7 density were found even in single boutons (Fig. 4*a*) when the postsynaptic targets were a dendrite of a presumed mGluR1 α -expressing interneuron and a pyramidal-cell dendritic spine. These results demonstrate that the distinct levels of presynaptic mGluR7 density depend on the identity of the postsynaptic target.

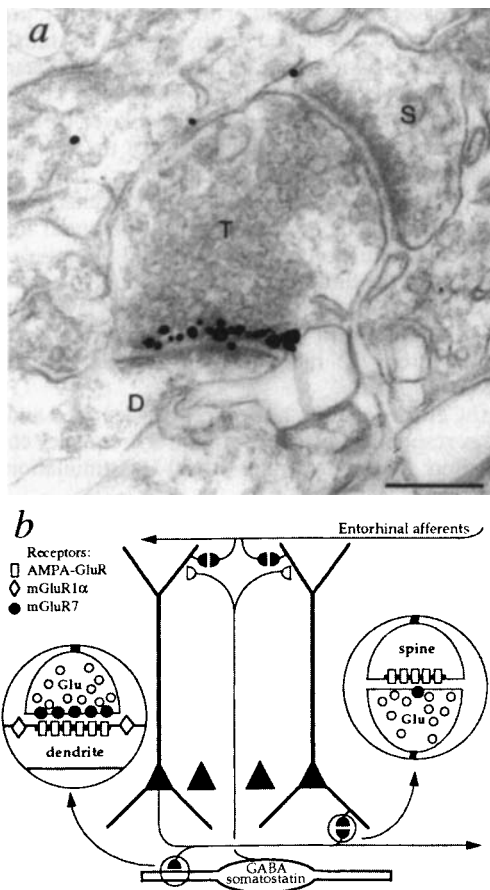


FIG. 4 Postsynaptic target determines presynaptic receptor density. *a*, The presynaptic receptor mGluR7 is differentially expressed at two synapses of a single nerve terminal (T), as demonstrated by immunogold labelling in the CA3 area. The heavily receptor immunolabelled presynaptic grid (particles) faces a dendritic shaft (D) characteristic of mGluR1 α -expressing interneurons, whereas the unlabelled synapse is on a pyramidal cell dendritic spine (S). Scale bar, 0.2 μ m. *b*, Summary of the molecular architecture of pyramidal-cell synapses demonstrating the position of some of the glutamate receptors in relation to the mGluR1 α /somatostatin/GABA-containing interneuron. Ionotropic α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA)-type receptors are enriched in the postsynaptic membrane specialization on both the dendritic spines of pyramidal cells and on the dendritic shafts of interneurons³¹. Postsynaptic mGluR1 α is concentrated in a perisynaptic annulus outside the synaptic junction¹⁰. The somatostatin/GABA-containing interneurons, which receive mainly recurrent axon collateral input^{10,24,25}, terminate in conjunction with the entorhinal cortical input on distal dendrites²⁷. There is a high density of the presynaptic autoreceptor mGluR7 in the presynaptic grid of pyramidal terminals which target mGluR1 α -positive cells, but only a low density in those terminals which innervate pyramidal-cell spines. This selective distribution may result in a low-pass frequency filter for glutamate release (see text).

The results provide evidence of a presynaptic receptor restricted to the site of transmitter release. Taken together with the low affinity of mGluR7 to glutamate (EC_{50} of 1 mM)⁸ compared with those of other mGluRs ($EC_{50} \leq 56 \mu$ M)³, it is possible that mGluR7 functions as an autoreceptor activated only by glutamate released at the site where the receptor is located. Such a synapse-specific autoregulation of transmitter release would be in contrast to the postulated heterosynaptic regulation mediated by mGluR2 (refs 12,13), which has much higher affinity to glutamate (EC_{50} of 12 μ M)³ than does mGluR7, and is located distant from the transmitter release site on presynaptic boutons and axons in the hippocampus¹¹. Voltage-sensitive calcium channels that trigger synaptic vesicle fusion¹⁴ are also thought to be concentrated at the presynaptic active zones¹⁵, and are inhibited

in a membrane-delimited manner by neurotransmitters through G-protein-coupled receptors^{14,16,17}. The mechanism of signal transduction for mGluR7 in the hippocampus is not known, but AP4-sensitive receptors pharmacologically similar to mGluR7 decrease the probability of glutamate release in the CA1 area^{5,6,18}. The apparently complete segregation of mGluR7 between two synapses within single boutons (Fig. 4a) suggests that coupling of the receptor with its effector is likely to be spatially restricted, and probably membrane delimited.

The clustering of postsynaptic receptors, regulated by trophic factors derived from the presynaptic nerve ending, has been extensively studied for nicotinic acetylcholine receptors at the neuromuscular junction¹⁹. In retinal bipolar cells, targeting of mGluR6 to postsynaptic sites is also dependent on the presynaptic neuronal element²⁰. The clustering of mGluR7 demonstrates a correlation between levels of presynaptic receptor expression and postsynaptic element identity. The phenomenon may underlie target-dependent variation in probability of transmitter release^{21,22}, and raises the possibility that postsynaptic neurons influence presynaptic receptor density in a retrograde manner.

That the input^{23,24} to mGluR1 α -expressing GABAergic^{10,25} cells is endowed by autoregulation stronger than that to pyramidal cells and other interneurons²⁶ might be due to their place in the hippocampal network (Fig. 4b). They make synapses in conjunction with the entorhinal input to pyramidal cells²⁷. The high level of presynaptic mGluR7 in the input terminals may suppress the release of glutamate when action potentials arrive at high frequency, allowing glutamate release to follow only relatively low-frequency presynaptic firing, that is, it could act as a low-pass filter. The activity of hippocampal principal cells and their entorhinal input shows gamma-frequency (30–60 Hz) oscillations modulated at theta (4–12 Hz) frequency^{28,29}. The time course of the recovery of glutamate release could be tuned to one of these frequencies, allowing the activation of the cells and recurrent GABA release to distal dendrites of pyramidal cells preferentially at one of the above frequencies. Thus the specifically high level of mGluR7 expression may provide pyramidal cells with a means to assist GABA-mediated timing of entorhinal input. □

Received 11 December 1995; accepted 4 April 1996.

- Nakanishi, S. & Masu, M. A. *Rev. Biophys. biomolec. Struct.* **23**, 319–348 (1994).
- Forsythe, I. D. & Clements, J. D. *J. Physiol., Lond.* **429**, 1–16, (1990).
- Pin, J.-P. & Duvoisin, R. *Neuropharmacology* **34**, 1–26 (1995).
- Trombley, P. Q. & Westbrook, G. L. *J. Neurosci.* **12**, 2043–2050 (1992).
- Manzoni, O. & Bockaert, J. *Eur. J. Neurosci.* **7**, 2518–2523 (1995).
- Gereau, R. W. & Conn, P. J. *J. Neurosci.* **15**, 6879–6889 (1995).
- Ohishi, H., Akazawa, C., Shigemoto, R., Nakanishi, S. & Mizuno, N. *J. comp. Neurol.* **360**, 555–570 (1995).
- Okamoto, N. et al. *J. Biol. Chem.* **269**, 1231–1236 (1994).
- Duvoisin, R. M., Zhang, C. X. & Ramonell, K. J. *Neurosci.* **15**, 3075–3083 (1995).
- Baude, A. et al. *Neuron* **11**, 771–787 (1993).
- Shigemoto, R. et al. *Soc. Neurosci. Abstr.* **21**, 846 (1995).
- Hayashi, Y. et al. *Nature* **366**, 687–690 (1993).
- Ohishi, H. et al. *Neuron* **13**, 55–66 (1994).
- Tareilus, E. & Breer, H. *Neurochem. Int.* **26**, 539–558 (1995).
- Ulinas, R., Sugimori, M. & Silver, R. B. *Science* **256**, 677–679 (1992).
- Clapham, D. E. A. *Rev. Neurosci.* **17**, 441–464 (1994).
- Sahara, Y. & Westbrook, G. L. *J. Neurosci.* **13**, 3041–3050 (1993).
- Vignes, M. et al. *Neuropharmacology* **34**, 973–982 (1995).
- Hall, Z. W. & Sanes, J. R. *Cell/Neuron* **72/10**, (suppl.) 99–121 (1993).
- Nomura, A. et al. *Cell* **77**, 361–369 (1994).
- Koerber, R. H. & Mendell, L. M. *J. Neurophysiol.* **65**, 590–597 (1991).
- Thomson, A. M. & Deuchars, J. *Trends Neurosci.* **17**, 119–126 (1994).
- Blasco-Ibanez, J. M. & Freund, T. F. *Eur. J. Neurosci.* **7**, 2170–2180 (1995).
- Maccaferri, G. & McBain, C. J. *Neuron* **15**, 137–145 (1995).
- Somogyi, P. et al. *J. Neurosci.* **4**, 2590–2603 (1984).
- Buhl, E. H., Halasy, K. & Somogyi, P. *Nature* **368**, 823–828 (1994).
- McBain, C. J., DiChiara, T. J. & Kauer, J. A. *J. Neurosci.* **14**, 4433–4445 (1994).
- Buzsaki, G., Leung, L.-W. & Vandervolf, C. H. *Brain Res. Rev.* **6**, 139–171 (1983).
- Chapack, S., Pare, D. & Ulinas, R. *Eur. J. Neurosci.* **7**, 1548–1557 (1995).
- Ohishi, H. et al. *Neurosci. Lett.* **202**, 85–88 (1995).
- Baude, A., Nusser, Z., Molnar, E., McIlhinney, R. A. J. & Somogyi, P. *Neuroscience* **69**, 1031–1055 (1995).

ACKNOWLEDGEMENTS. We thank E. Molnár for help in immunoblotting; A. D. Smith for comments on the manuscript; D. Latawiec for technical assistance; and P. Jays and F. Kennedy for photographic assistance. This work was partly supported by the Ministry of Education, Science and Culture of Japan. A.K. is supported by the MHB Magyar Tudományért Foundation and the OTKA Foundation of the Hungarian Government.

CORRESPONDENCE and requests for materials should be addressed to P.S.

Experience-dependent modification of synaptic plasticity in visual cortex

Alfredo Kirkwood, Marc G. Rioult & Mark F. Bear

Department of Neuroscience and Howard Hughes Medical Institute, Brown University, Providence, Rhode Island 02912, USA

In many regions of the cerebral cortex, Ca^{2+} influx through NMDA (*N*-methyl-D-aspartate) sensitive glutamate receptors (NMDA receptors) can trigger two forms of synaptic plasticity: long-term depression (LTD) and long-term potentiation (LTP)¹. LTD is induced by low levels of postsynaptic NMDA-receptor activation, for instance in response to low-frequency stimulation, whereas LTP is induced by the stronger activation that occurs following high-frequency stimulation²⁻⁴. Theoretical studies have shown that the properties of synaptic LTD and LTP can account for many aspects of experience-dependent plasticity in the developing visual cortex, provided that the LTD-LTP crossover point (the modification threshold, θ_m) varies as a function of the history of cortical activity⁵⁻⁷. Here we provide direct experimental evidence that the value of θ_m depends on sensory experience. We find in visual cortex of light-deprived rats that LTP is enhanced and LTD diminished over a range of stimulation frequencies, and that these effects can be reversed by as little as two days of light exposure. Our findings support the idea that a variable synaptic-modification threshold allows synaptic weights in neural networks to achieve a stable equilibrium.

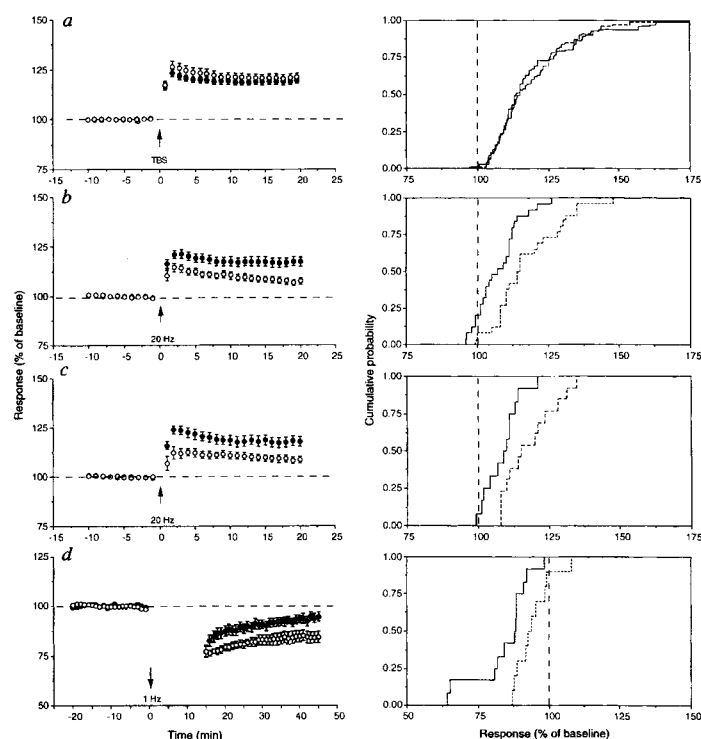
The Bienenstock-Cooper-Munro (BCM) theory was originally proposed to account for aspects of experience-dependent visual-cortical plasticity⁶. It assumes that active synapses undergo LTD or LTP depending on the level of postsynaptic response, an assumption for which there is now good evidence both in hippocampus and neocortex^{2-4,8-11}. It also assumes that the value of the threshold θ_m is not fixed, but varies as a function of the previous activity of the postsynaptic cortical neuron. Thus, θ_m increases after a period of increased activity, promoting synaptic depres-

sion, and decreases after a period of decreased activity, promoting synaptic potentiation. We tested this hypothesis by studying LTD and LTP of layer III synaptic responses in slices of visual cortex prepared from light-deprived and control rats 4-6 weeks old.

The field potentials evoked in layer III by layer IV stimulation were of similar size and shape, regardless of the rearing history (light-deprived or not) of the visual cortex. There were no significant differences in the stimulation currents required to evoke half-maximal field potentials, the field-potential amplitudes at half-maximal stimulation intensity, the field-potential widths at half-amplitude, or the times to peak. In addition, a comparison of LTP induced by theta-burst stimulation (TBS) of layer IV in visual cortex of light-deprived ($n = 68$ slices from 29 rats) and control ($n = 70$ slices from 32 rats) animals revealed no significant difference (Fig. 1a).

Unlike the effects of TBS, however, light-deprived and control visual cortex responded differently to lower-frequency conditioning stimulation. Three 2-s trains of 20-Hz stimulation, which

FIG. 1 Activity-dependent modification of synaptic responses in visual-cortical slices from light-deprived (filled symbols) and normal (open symbols) rats. Left, average ($\pm$ s.e.m.) responses (amplitude of the maximum negative field potential in layer III normalized to average baseline value) and their modification by tetanic stimulation (arrows). Right, cumulative histograms showing the effects of conditioning stimulation in every slice from both groups (light-deprived, dotted lines; control, solid line). a, Effects of three trains of TBS (120 pulses total). b, Effects of three trains of 20-Hz stimulation (120 pulses total) in all cases studied. c, Effects of 20-Hz stimulation in experiments in which the investigator was 'blind' to the rearing history. d, Effects of 1-Hz stimulation for 15 min (900 pulses total). METHODS. Dark rearing and brain slice preparation were performed as described¹⁵. Slices of visual cortex were maintained in humidified 95% O_2 , 5% CO_2 , and superfused with 30°C artificial cerebrospinal fluid (ACSF) at a rate of 1 ml min⁻¹. The ACSF was saturated with 95% O_2 , 5% CO_2 , and contained (in mM) NaCl, 124, KCl 5, NaH_2PO_4 1.25, MgSO_4 1, CaCl_2 2, NaHCO_3 26, dextrose 10. A site in the middle of the cortical thickness, confirmed histologically to correspond with layer IV and upper layer V, was stimulated to evoke field potentials in layer III, as described^{15,29}. The amplitude of the maximum negative field potential in layer III was used as a measure of the evoked population excitatory synaptic current. Changes in the amplitude of the maximum negative field potential reflect changes in the magnitude of a monosynaptic current sink, and correlate with changes in the initial slope of excitatory postsynaptic potentials recorded intracellularly in layer III neurons^{3,29}. Baseline responses were obtained every 15 s with a stimulation intensity that yielded a half-maximal response. To study the stimulation requirements for inducing LTP, stimulus trains 2 s long were repeated every 10 s until 120 pulses had been delivered. Typically, either three 20-Hz trains or three trains of TBS were delivered; a train of TBS consists of brief bursts of stimuli delivered every 200 ms, with each burst containing 4 pulses at 100 Hz. In some experiments, six 10-Hz trains were also used. To study LTD, 900 pulses were delivered at 1 or 2 Hz. The stimulation pulse duration and intensity during conditioning stimulation were the same as for baseline stimulation. Induction of LTP and LTD with these stimulation protocols requires NMDA-receptor activation^{3,29}. The group data were analysed as follows: the maximum negative field-potential amplitude data for each experiment were expressed as percentages of the preconditioning baseline average; the timescale in each experiment was converted to time from the onset of conditioning; and the time-matched, normalized data were averaged across experiments and expressed as the means ($\pm$ s.e.m.). Light-deprived and control groups were compared 20 min after cessation of HFS, and 30 min after LFS, using a *t*-test. Cumulative histograms were also constructed to show the data from each slice in each experimental group, as described¹⁵. For the 'blind' experiments, one light-deprived and one control animal was used each day, studied simultaneously on two slice rigs. For each slice, LTP was first attempted with TBS, which previous experiments had shown to be unaffected (on average) by rearing history or age¹⁵. If plasticity was observed, LTP was attempted using 20-Hz trains at a distant, independent location on the same slice; if no plasticity was observed, the slice was not studied further. We used this criterion to ensure that only good-quality slices were used in the 'blind' study. The only criterion used for the non-blind study, however, was a stable baseline. The gross morphology of slices from light-deprived and control animals was indistinguishable³⁰.



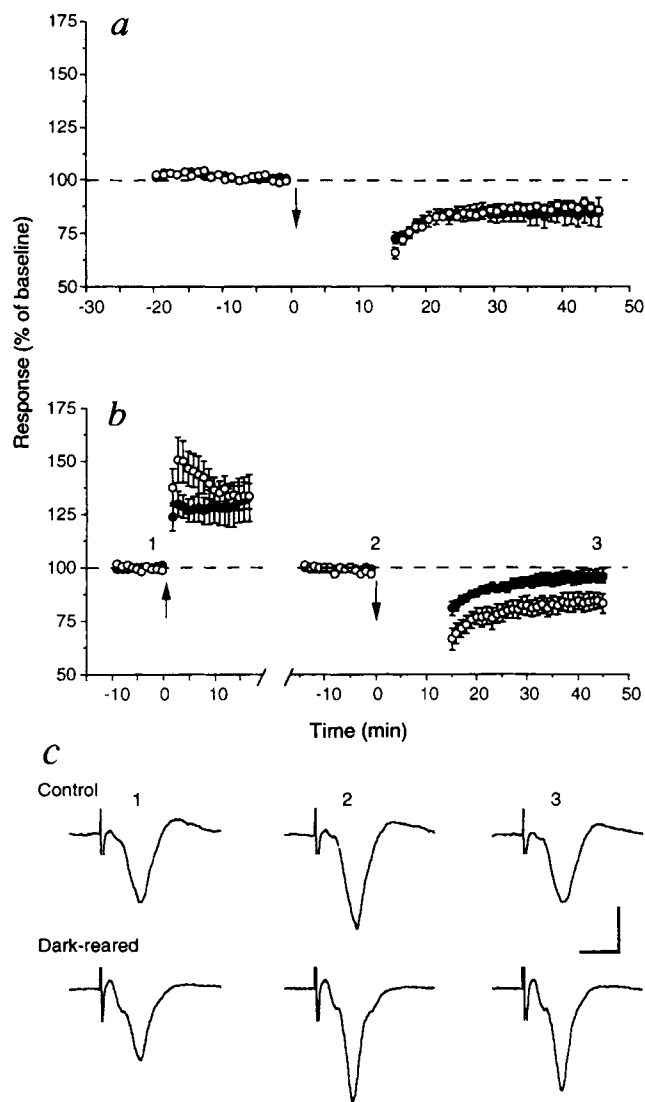


FIG. 2 **a**, Investigation of LTD induction in area CA1 of hippocampus in slices from light-deprived (filled symbols) and normal (open symbols) rats. **b**, Investigation of LTD induction following establishment of LTP in visual cortex of light-deprived and normal rats. LTP was induced with TBS, LTD with LFS (1 Hz for 15 min). Time elapsed between TBS and LFS varied from 25 to 45 min. Data shown to the right of the break in the x-axis were renormalized. **c**, Representative field potentials evoked in layer III by layer IV stimulation in visual cortex of light-deprived and control rats. Each trace is the average of 4 consecutive responses, collected at the times indicated by the numbers in **b**. Scale bars: 5 ms, 0.8 mV for control; 5 ms, 1.0 mV for light-deprived rats.

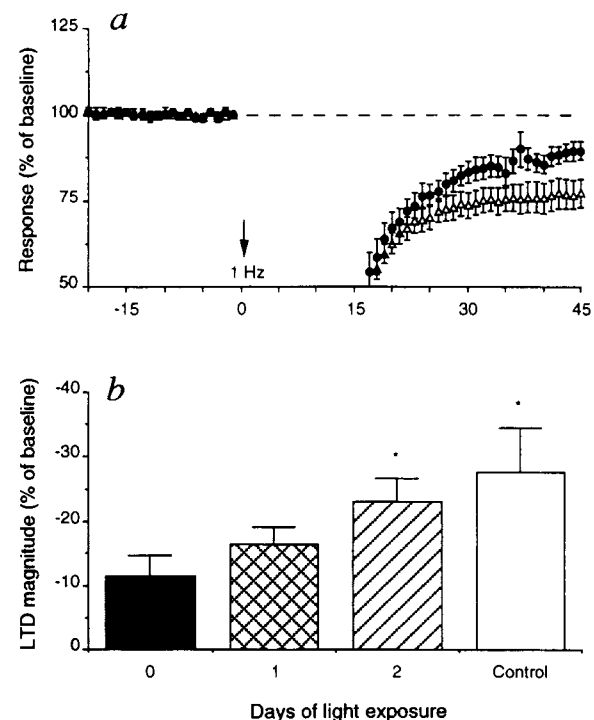


FIG. 3 Visual experience restores normal LTD to light-deprived visual cortex. **a**, The effects of 1-Hz stimulation (900 pulses) on synaptic responses in light-deprived cortex (filled symbols) and in visually deprived cortex receiving 2 days of light exposure before slice preparation (open symbols). **b**, Comparison of the magnitude of LTD 30 min after LFS in visual cortex of animals receiving various amounts of light exposure. Both the 2-day light-exposure and control groups significantly differ from light-deprived (asterisk, Tukey test, $P < 0.05$).

METHODS. Visually deprived rats 5–6 weeks old were killed 24 or 48 h after light exposure. Slices were prepared as described (Fig. 1) and, in these experiments, were maintained in a submersion chamber at 31 °C. The composition and flow rate of the ACSF was as Fig. 1. Data were collected, analysed and displayed as Fig. 1. All LFS cases were pooled, even those following application of a high-frequency tetanus, as the magnitude of LTD was independent of whether LTP had been previously induced or not (Figs 1d and 2b).

Visually evoked responses in light-deprived visual cortex are weak and unreliable¹². The reduced LTD in slices from light-deprived animals could therefore be due to synaptic strengths already being near the 'floor' of their available dynamic range. To test this, we took advantage of the fact that the effectiveness of

deliver the same number of pulses over the same time interval as TBS, produced stable LTP in slices from light-deprived animals ($117.4 \pm 2.3\%$ of baseline, $n = 26$ slices from 17 rats), but little potentiation in controls ($107.3 \pm 1.5\%$, $n = 25$ slices from 18 rats; Fig. 1b). The difference between light-deprived and control groups was significant at $P < 0.005$. To rule out any possible contribution of experimenter bias to the results, the effects of 20-Hz stimulation were re-examined in a series of experiments in which the investigator was 'blind' to the rearing history of the rats. These studies yielded essentially identical results ($117.8 \pm 2.5\%$, $n = 13$ slices from 9 light-deprived rats, compared to $108.6 \pm 1.8\%$, $n = 12$ slices from 9 control rats; $P < 0.01$; Fig. 1c). Potentiation following 10-Hz trains was also of greater magnitude in slices from light-deprived rats ($115.0 \pm 5.8\%$, $n = 5$ slices from 4 rats) compared with control ($105.0 \pm 2.7\%$, $n = 9$ slices from 8 rats; $P < 0.07$; data not shown).

Induction of homosynaptic LTD was studied by delivering 900 pulses at 1 Hz. In slices from control animals, low frequency stimulation (LFS) produced significant LTD ($84.3 \pm 2.4\%$, $n = 12$ slices from 10 rats). In light-deprived animals, however, the magnitude of depression after LFS was significantly less ($94.4 \pm 2.3\%$, $n = 9$ slices from 7 rats; $P < 0.02$; Fig. 1d). The difference in LTD was regionally specific; there was no significant difference in the magnitude of hippocampal LTD in slices from light-deprived ($n = 4$ from 4 rats) and control ($n = 5$ slices from 5 rats) animals (Fig. 2a).

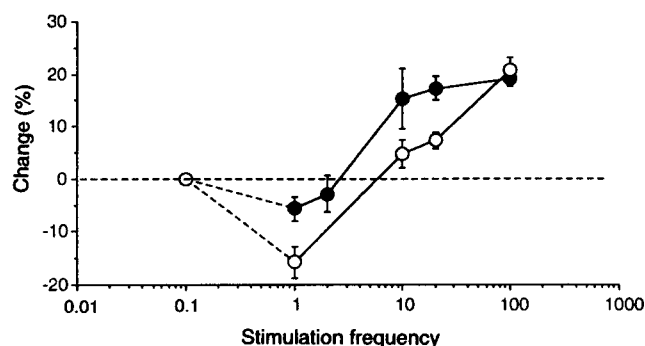


FIG. 4 Frequency-response functions derived from visual cortex of light-deprived (filled symbols) and normal (open symbols) rats. Data points for stimulation frequencies ≥ 10 Hz represent the average change ($\pm$ s.e.m.) 20 min after the delivery of 120 pulses of conditioning stimulation. Data points for 1- and 2-Hz stimulation represent the average change ($\pm$ s.e.m.) 30 min after delivery of 900 pulses of conditioning stimulation. The data point for 0.07 Hz is inferred as baseline stimulation once every 15 s does not appear to induce synaptic modification in light-deprived or normal cortex. All data are from slices maintained in an interface chamber (see Fig. 1).

individual synapses can be bidirectionally modified by high- and low-frequency stimulation^{8,13}. Thus, in an additional series of experiments, synaptic strengths were first raised by inducing LTP with TBS, and then, after waiting for a stable baseline to be reestablished (25–45 min), 1-Hz stimulation was delivered. LFS was still significantly less effective in producing synaptic depression in visual cortex from light-deprived animals (Fig. 2b, c). The response 30 min after 1-Hz stimulation, measured as a percentage of the stable pre-LFS baseline, was $83.3 \pm 4.5\%$ in control slices ($n = 6$ slices from 4 rats), but only $96.0 \pm 3.2\%$ in slices from light-deprived animals ($n = 6$ slices from 4 rats; $P < 0.05$). Together, the data strongly suggest that the LTD-induction mechanism, or its recruitment by 1-Hz stimulation, is altered in light-deprived cortex.

Total light deprivation can slow many aspects of visual cortical development^{12,14,15}, so it is possible that the altered frequency-response function in light-deprived cortex simply reflects an immature state. We consider this explanation unlikely, however. Both in hippocampus^{13,16,17} and in visual cortex (unpublished observations), LFS produces significantly greater LTD in neonatal animals than in young adults. In contrast, dark rearing for several weeks appears to result in diminished LTD. A more likely explanation is that the frequency-response function has shifted during postnatal development as a specific consequence of cortical inactivity.

As an additional test of the sliding- θ_m hypothesis, visually deprived rats were exposed to light for various times, and the effects of LFS were investigated in visual cortex. In these experiments, the response 30 min after 1-Hz stimulation was $88.6 \pm 3.2\%$ of baseline in slices from light-deprived animals ($n = 15$ slices from 7 rats). The slightly greater average LTD magnitude in this group might be due to the slices being maintained in a submersion chamber instead of the interface chamber used in the previous experiments. In any case, LTD in slices from light-deprived animals was still significantly less than in controls studied under identical conditions ($72.4 \pm 6.8\%$; $n = 5$ slices from 5 rats; $P < 0.02$). Remarkably, however, the magnitude of LTD in light-deprived visual cortex returned nearly to control levels after only 2 days of light exposure (Fig. 3). The response 30 min after LFS was $77.0 \pm 3.6\%$ of baseline in cortex exposed to light for 2 days ($n = 11$ slices from 5 rats), which is significantly greater than in light-deprived cortex ($P < 0.02$). These data are consistent with the hypothesis that θ_m 'slides' as average cortical activity increases.

The time course of the observed change in LTD closely corresponds to that predicted for θ_m in modelling studies¹⁸.

Activity-dependent regulation of NMDA-receptor-dependent synaptic plasticity can occur at many levels, ranging from changes in network inhibition to alterations in postsynaptic NMDA receptors and Ca^{2+} -binding proteins^{4,19–23}. Our data do not directly address which mechanism(s) accounts for the shift of the frequency-response function in light-deprived visual cortex. It is also unclear to what extent the present findings relate to the increased susceptibility to LTD that occurs transiently following synaptic stimulation in hippocampus *in vitro*^{4,24}. Indeed, in visual cortex the magnitude of LTD caused by a single episode of LFS is the same regardless of whether LTP had previously been induced (Figs 1d and 2b).

Nonetheless, at a macroscopic level, synaptic plasticity is altered by light deprivation such that LTP is promoted over LTD over a range of stimulation frequencies (Fig. 4). The consequences of such an alteration on visual-cortical development have been investigated in theoretical studies^{18,25}. For example, enhanced LTP enables initially weak visual responses to undergo rapid potentiation on subsequent light exposure, a trait characteristic of visual cortex in light-deprived animals^{26,27}.

Our findings lend experimental support to the concept that θ_m is not fixed, but varies according to the activation history of the cortex. Thus, key assumptions of the BCM theory have now been shown to have a plausible physiological basis. Analysis and computer simulations have shown that the BCM theory can account for diverse modifications of visual-cortical receptive fields that follow various manipulations of visual experience^{18,25}. Our findings therefore provide additional support for the unifying hypothesis that the mechanisms of LTD and LTP account for key aspects of experience-dependent synaptic modifications in the visual cortex⁷. More generally, our results demonstrate that the properties of activity-dependent synaptic plasticity are themselves dependent on experience. Such plasticity of synaptic plasticity, or metaplasticity²⁸, may be critical for information storage by synapses in the brain. □

Received 20 February; accepted 19 April 1996.

1. Bear, M. F. & Malenka, R. C. *Curr. Opin. Neurobiol.* **4**, 389–399 (1994).
2. Dudek, S. M. & Bear, M. F. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4363–4367 (1992).
3. Kirkwood, A. & Bear, M. F. *J. Neurosci.* **14**, 3404–3412 (1994).
4. Mayford, M., Wang, J., Kandel, E. & O'Dell, T. *Cell* **81**, 1–20 (1995).
5. Sejnowski, T. J. *J. math. Biol.* **69**, 385–389 (1977).
6. Bienenstock, E. L., Cooper, L. N. & Munro, P. W. *J. Neurosci.* **2**, 32–48 (1982).
7. Bear, M. F., Cooper, L. N. & Ebner, F. F. *Science* **237**, 42–48 (1987).
8. Mulkey, R. M. & Malenka, R. C. *Soc. Neurosci. Abstr.* **18**, 1498 (1992).
9. Artola, A., Brocher, S. & Singer, W. *Nature* **347**, 69–72 (1990).
10. Kimura, F., Tsumoto, T., Nishigori, A. & Yoshimura, Y. *Neuroreport* **1**, 65–68 (1990).
11. Frégnac, Y., Burke, J., Smith, D. & Friedlander, M. *J. Neurophysiol.* **71**, 1403–1421 (1994).
12. Fagioli, M., Pizzorusso, T., Berardi, N., Domenici, L. & Maffei, L. *Vision Res.* **34**, 709–720 (1994).
13. Dudek, S. M. & Bear, M. F. *J. Neurosci.* **13**, 2910–2918 (1993).
14. Cynader, M. & Mitchell, D. E. *J. Neurophysiol.* **43**, 1026–1039 (1980).
15. Kirkwood, A., Lee, H.-K. & Bear, M. F. *Nature* **375**, 328–331 (1995).
16. Mulkey, R. M. & Malenka, R. C. *Neuron* **9**, 967–975 (1992).
17. Wagner, J. J. & Alger, B. E. *J. Neurosci.* **15**, 1577–1586 (1995).
18. Clothiaux, E. E., Bear, M. F. & Cooper, L. N. *J. Neurophysiol.* **66**, 1785–1804 (1991).
19. Steele, P. M. & Mauk, M. D. *Soc. Neurosci. Abstr.* **20**, 1514 (1994).
20. Kerr, D. S. & Abraham, W. C. *Proc. natn. Acad. Sci. U.S.A.* **62**, 11637–11641 (1996).
21. Fox, K., Daw, N., Sato, H. & Czeplita, D. *Nature* **350**, 342–344 (1991).
22. Carmignoto, G. & Vicini, S. *Science* **258**, 1007–1011 (1992).
23. Selig, D., Hjelmstad, G., Herron, C., Nicoll, R. & Malenka, R. *Neuron* **15**, 417–426 (1995).
24. Wexler, E. M. & Stanton, P. R. *NeuroReport* **4**, 591–594 (1993).
25. Shouval, H., Intrator, N., Law, C. C. & Cooper, L. N. *Neural Comput.* (in the press).
26. Buisseret, P. & Imbert, M. *J. Physiol. Lond.* **255**, 511–525 (1976).
27. Buisseret, P., Gary-Bobo, E. & Imbert, M. *Nature* **272**, 816–817 (1978).
28. Abraham, W. C. & Bear, M. F. *Trends Neurosci.* **19**, 126–129 (1996).
29. Kirkwood, A. & Bear, M. F. *J. Neurosci.* **14**, 1634–1645 (1994).
30. Gabbott, P., Stewart, M. & Rose, S. *Expl. Brain Res.* **64**, 225–232 (1986).

ACKNOWLEDGEMENTS. We thank A. Heynen and K. Huber for comments on the manuscript, and L. N. Cooper for discussions. This work was supported by the Charles A. Dana Foundation and the National Science Foundation.

CORRESPONDENCE and requests for materials should be addressed to M.F.B. (e-mail: Mark_Bear@brown.edu).

Genetic modification of heterochromatic association and nuclear organization in *Drosophila*

Amy K. Csink & Steven Henikoff

Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104, USA

HETEROCHROMATIN is the highly compact, usually pericentromeric, region of eukaryotic chromosomes. Unlike the more gene-rich euchromatin, heterochromatin remains condensed during interphase, when it is sequestered to the periphery of the nucleus¹⁻³. Here we show, by using fluorescent *in situ* hybridization to interphase diploid nuclei of *Drosophila*, that the insertion of heterochromatin into a euchromatic gene, which results in position-effect variegation (PEV), also causes the aberrant association of the gene and its homologous copy with heterochromatin. In correlation with the gene's mutant variegating phenotype, the cytological association of the heterochromatic region is affected by chromosomal distance from heterochromatin and by genic modifiers of PEV. Proteins that are thought to be involved in the formation of heterochromatin can therefore influence the interphase nuclear position of a chromosomal region. This suggests that heterochromatin and proteins involved in its formation provide a structural framework for the interphase nucleus.

The *brown* (*bw*⁺) eye-colour locus of *Drosophila melanogaster* is located near the distal end of the right arm of chromosome 2 (2R) at polytene band 59E. An insertion of a large block (>1 megabase) of heterochromatin containing AAGAG satellite into the coding region of *brown* has resulted in the *brown*^{Dominant} (*bw*^D) allele of this gene⁴ (P. Dimitri, personal communication). In *bw*^D/*bw*⁺ heterozygous flies, this insertion acts *in trans* to 'trans-inactivate' the *bw*⁺ copy of *brown*, resulting in a variegated eye with only 2% of *bw*⁺ expression and no detectable messenger RNA accumulation⁵. Trans-inactivation is typical for PEV alleles of *brown*, but is rare for other genes. Previous work demonstrated a *bw*^D-heterochromatin distance effect on *bw*^D/*bw*⁺ heterozygotes⁶. Chromosomal rearrangements that move the *bw*^D allele further from centric heterochromatin suppress its trans-inactivation ability, whereas those that move *bw*^D nearer enhance trans-inactivation. In such distance-enhanced lines, the *bw*^D locus associates more frequently with the chromocentre in polytene salivary-gland nuclei⁶.

To determine whether the *bw*^D-heterochromatin association observed in polytene chromosomes also occurs in diploid cells, we used fluorescent *in situ* hybridization (FISH) painting of the interphase chromosomes of larval neuroblasts. Two probes were used: one for AACAC, a satellite sequence specific to the centric heterochromatin of 2R (2Rh)⁷, and a second probe consisting of a unique genomic clone of ~80 kilobases (kb) just proximal to the *brown* gene (59E). As well as positions of each of the above two hybridization signals on condensed mitotic chromosome spreads, we also show the positions of the AAGAG satellite (Fig. 1). Note the high concentration of AAGAG sequences in 2Rh, as well as its unique presence on the tip of 2R in *bw*^D. In the interphase nuclei (Fig. 1), the signals from the 59E (red) and AACAC (cyan) probes are closer to each other in the *bw*^D nuclei than in the wild-type nuclei. To confirm this observation, data were collected and analysed from 90–180 nuclei from each genotype. Distance between each hybridization signal (red and green in Fig. 2a) was measured and each measurement was divided by the nuclear radius. The median distance between distal 59E and 2Rh decreases in *bw*^D neuroblasts when compared to *bw*⁺, which controls for chance coincidence of the two signals (Fig. 2b). Further, the distribution of the distances changes (Fig. 2c).

Unlike the distribution seen for *bw*⁺, the distribution for *bw*^D is strongly skewed towards zero. Perhaps this is because in *bw*⁺ the two sequences are unconstrained, but in a subset of *bw*^D nuclei distal 2R adheres to 2Rh, while in the remainder of the *bw*^D nuclei the two chromosomal regions are unconstrained. This would result in two superimposed distributions in the *bw*^D data set.

To determine whether association between 2Rh and 59E correlates with *bw*^D trans-inactivation, we first analysed *bw*^D/*bw*⁺ heterozygotes. The level of association between 2Rh and 59E in *bw*^D/*bw*⁺ heterozygotes was not significantly different from that in *bw*^D homozygotes (Fig. 2b). This implies that trans-inactivation is due to the pairing of the homologues and the consequent dragging

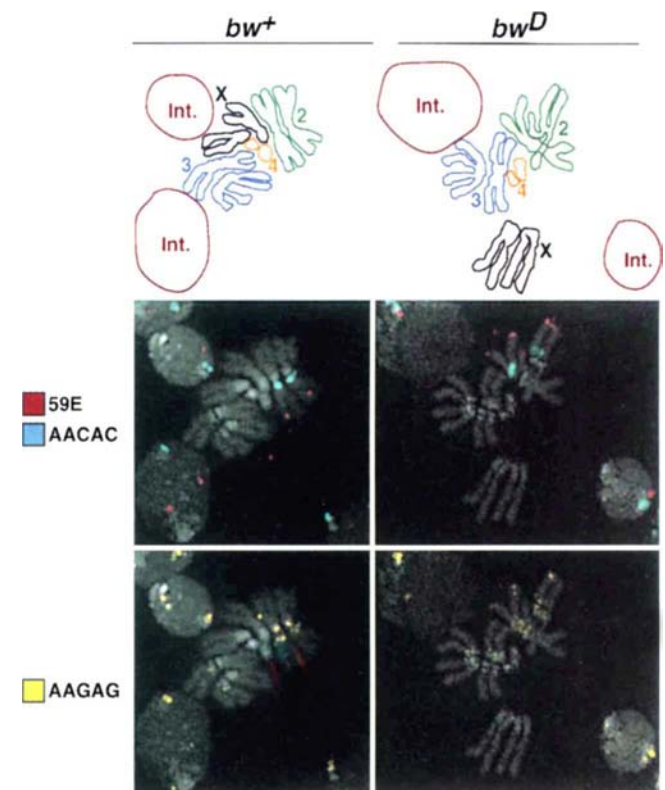
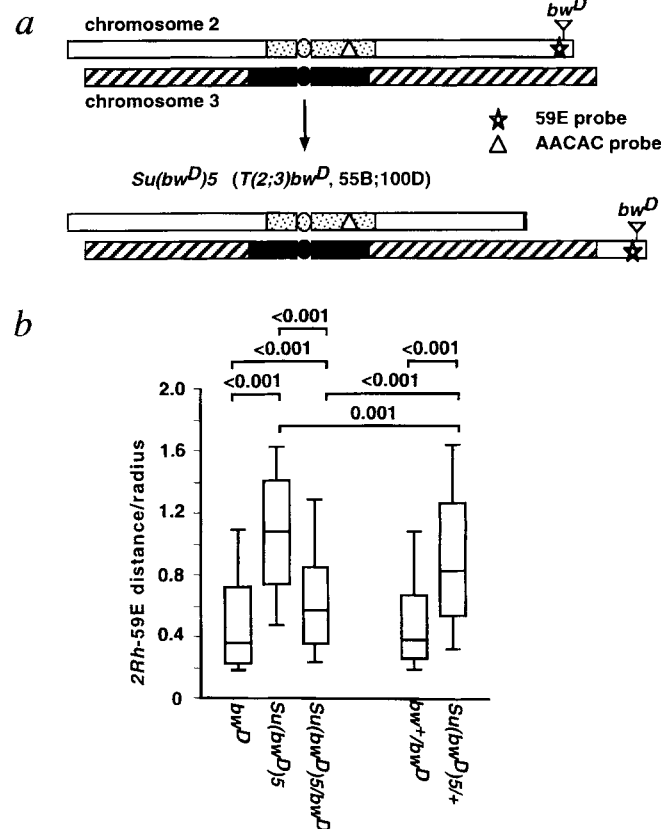
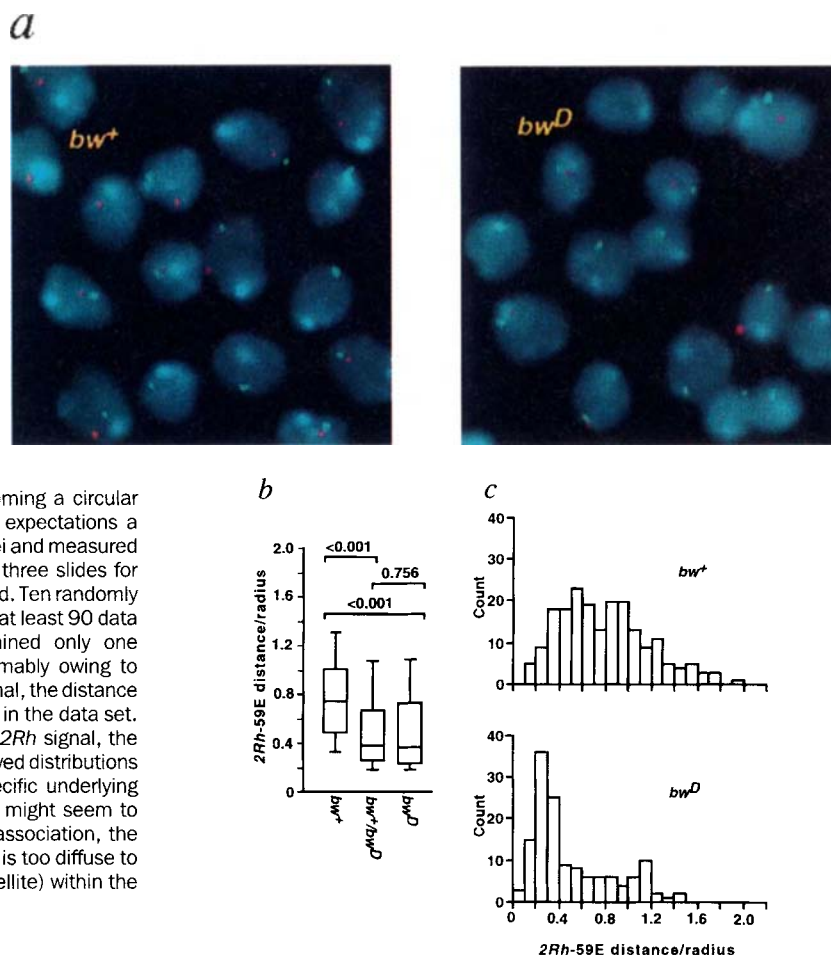


FIG. 1 *Drosophila* neuroblasts from wild-type (*bw*⁺) and *bw*^D female larvae showing both mitotic and interphase nuclei hybridized to three separately labelled probes. Each of the four chromosomes is indicated in the upper panels. Int., interphase.

METHODS. Brains were dissected from late third-instar female larvae and prepared according to ref. 19 but without colchicine, and tissue was treated with 1% sodium citrate for 5 min. Nuclei were hybridized according to the method for biotinylated DNA *in situ* hybridization¹⁹ with the following modifications. Slides were treated with 200 µg ml⁻¹ RNase in 2× SSC for 1 h at 37 °C. Each slide was hybridized with 12 µl of 50% formamide, 10% dextran sulphate, and 2× SSC with 20–40 ng of each probe for 16–24 h at 37 °C. Post-hybridization washes consisted of 2 changes of 50% formamide, 2× SSC at 37 °C for 30 min and three PBS washes for 5 min. Slides were then treated with anti-digoxigenin-fluorescein, Fab fragments (Boehringer Mannheim), and Cy5-conjugated streptavidin (Jackson ImmunoResearch) in PBS and 1% BSA for 1 h at room temperature, stained with 0.2 µg ml⁻¹ DAPI in water for 20 s, air dried and mounted in Vectashield (Vector). Oligonucleotide (40–50 bases) probes for the satellites AACAC and AAGAG were end-labelled using terminal deoxynucleotidyl transferase (GibcoBRL) and biotin-16-dUTP and digoxigenin-11-dUTP, respectively, according to manufacturer's instructions. The 59E probe is a genomic clone in a P1 viral vector provided by the *Drosophila* Genome Project (P1 37–24)²⁰. P1 DNA was isolated and cut with a mixture of 4 bp restriction enzymes to yield an average size of fragments between 50 and 150 bp, which were then end-labelled with rhodamine-4-dUTP (Amersham)²¹. Nuclei were visualized using a Deltavision SA3 microscope (Applied Precision) with a cooled CCD camera; magnification, ×1,500. Each of the four wavelengths was corrected using the Deltavision 2D deconvolution program (Applied Precision) and merged using Adobe Photoshop.

FIG. 2 a, Example of images from which the accompanying data were collected. DAPI, blue; AACAC, green; 59E, red. b, Box plots of the distance between the 2Rh probe (AACAC) and the 59E probe divided by the radius of the nucleus. Box plots are calibrated representations of histograms in which each horizontal line represents the 10th, 25th, 50th (median), 75th and 90th percentiles; genotypes are indicated below. The numbers on the brackets indicated the *P*-value for the bracketed pair as determined using the two-tailed Mann-Whitney *U* test for non-parametric comparison of two unpaired groups²². c, Histogram of data from the *bw*⁺ and *bw*^D genotypes. METHODS. Neuroblasts from the brains of late third-instar female larvae were hybridized as in Fig. 1, and images collected at an original magnification of $\times 600$. The image for each wavelength from the CCD camera were combined in Photoshop, and NIH Image (W. Rasband) was used to measure the distance between the probes (red and green signals in a) and the area of the DAPI-stained nuclei (blue in a). The radius of each nucleus was calculated based on the area, assuming a circular shape. To ensure that measurements were unbiased by expectations a student with no training in genetics randomly selected nuclei and measured nuclear area and signal distances. From each of at least three slides for each genotype, three separate fields were randomly selected. Ten randomly selected nuclei were analysed from each field, for a total of at least 90 data points for each genotype. Most (>75%) nuclei contained only one hybridization signal from each of the two probes, presumably owing to somatic pairing. In nuclei containing more than one 59E signal, the distance between each 59E signal and the 2Rh signal was included in the data set. For internal consistency, where there was more than one 2Rh signal, the closest one to the 59E signal was used. Because of the skewed distributions we applied a rank test, which does not assume any specific underlying distribution, to determine a *P*-value. Although these data might seem to allow for an estimation of the fraction of nuclei showing association, the AAGAG labelling (Fig. 1) suggests that 2R heterochromatin is too diffuse to model association based on a single point (the AACAC satellite) within the cloud.



of the wild-type homologue by the *bw*^D heterochromatic element into association with 2Rh.

Because the variegated phenotype of *bw*^D accompanies the localization of sequences to a distinct region of the nucleus, we investigated the possibility that various genetic modifications of the *bw*^D phenotype would also modify this nuclear localization. To see if phenotypic modification correlates with nuclear localization, we used a translocation of distal 2R containing *bw*^D to the tip of 3R (Fig. 3a). This translocation is one of 75 rearrangements in which the distance of *bw*^D from other heterochromatin is inversely correlated (without exception) with trans-inactivation^{4,6}. When this chromosome, *Su(bw*^D*)5*, is heterozygous with a *bw*⁺ chromosome, *bw*^D trans-inactivation is partly suppressed by distance. Here we show that association between 2Rh and 59E is also suppressed, both for homozygotes (Fig. 3b, left) and for heterozygotes (right). The data also suggest that the pairing of the translocated end of 2R (now attached to 3R) with an unrearranged chromosome can drag 59E into closer association with 2Rh. Relative to the homozygous translocation *Su(bw*^D*)5*, 2Rh association is facilitated by the presence of an unrearranged *bw*⁺ chromosome [*Su(bw*^D*)5/+*] or *bw*^D chromosome [*Su(bw*^D*)5/bw*^D].

PEV alleles of many *Drosophila* genes are generated by chromosomal rearrangements that lead to the abnormal proximity of heterochromatin and euchromatin (see ref. 8 for review). Nearly all of these PEV alleles are dominantly modified by *Suppressor-of-variegation* (*Su(var)*) mutations. Only *Su(var)205* is known to be a structural component of heterochromatin, so we chose this, as well as two additional *Su(var)*s (3-8 and 3-9). All three of these *Su(var)*s suppress *bw*^D trans-inactivation, as do other *Su(var)*s

FIG. 3 a, The derivation of the distance-suppressed *Su(bw*^D*)5*⁶ translocation from the *bw*^D chromosome 2 and a normal chromosome 3. b, Box plots of the *Su(bw*^D*)5* data set.

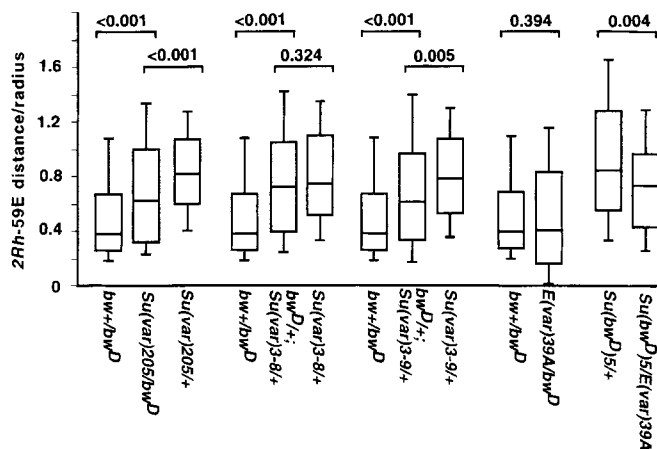


FIG. 4 Box plots showing the effects of modifiers of PEV on association between 2Rh and 59E.

METHODS. To produce larvae of the genotypes shown, males of the genotypes²³ *Su(var)205*⁵⁰²/*ln(2LR)Gla, Bc*, *E(var)39A/ln(2LR)Gla, Bc*, *Su(var)3-8/TM6, Tb* or *Su(var)3-9/TM6, Tb* were crossed to either wild-type, *bw*^D or *Su(bw)*⁵ virgin females, and the *Bc*⁺ or *Tb*⁺ larvae were chosen for hybridization.

(G. Sass and S.H., unpublished data). On testing these three *Su(var)*s, we found that association of 2Rh and 59E in larval neuroblasts was also dominantly suppressed (Fig. 4). One of the genes, *Su(var)205*, encodes heterochromatin protein (HP)-1, which predominantly associates with heterochromatin in both polytene⁹ and diploid nuclei¹⁰. When HP-1 is present in three doses, as when the duplication allele *E(var)39A* is present, PEV is enhanced¹¹. In the presence of this enhancer, the *Su(bw)*⁵ chromosome shows increased association between 2Rh and 59E (Fig. 4).

Unexpectedly, a standard *bw*^D chromosome does not show enhanced association between 2Rh and 59E when *E(var)39A* is present. All of our data are collected on larval brain tissue, which is a heterogeneous collection of cells, certainly at different stages in the cell cycle, and perhaps at different levels of differentiation. In the eye, *bw*^D/*bw*⁺ expresses only 2% of the wild-type brown pigment, and a decrease of 2% to zero may well be undetectable by our experimental design. However, the tail on the histogram (Fig. 2c), as seen persistently in the box plots, suggests that a substantial fraction of nuclei in the *bw*^D/*bw*⁺ and *bw*^D data sets can be interpreted as showing no association. This observation, and the result that this non-associating subpopulation was refractory to the effects of *E(var)39A*, suggests that other factors, such as stage of the cell cycle or time since last cell division, influence the formation of heterochromatin. Indeed, lack of association between 2Rh and 59E at metaphase (Fig. 1) is consistently observed for all genotypes. Perhaps this is related to the absence of detectable HP-1 binding during mitosis¹⁰; it has been reported that nuclear architecture varies with the cell cycle, tissue type, and level of differentiation¹²⁻¹⁴.

An effect of specific nuclear position on gene expression has been the motivation for many studies of nuclear architecture¹⁵⁻¹⁷, but evidence of differential phenotypes caused by nuclear mislocalization in diploid cells has been lacking¹⁸. Our results show that decreased expression of the *brown* gene in eye tissue accompanies its sequestration to the heterochromatic compartment of the interphase nucleus of neuroblasts, and suggest that *Su(var)* proteins are responsible for this sequestration.

Note added in proof: A similar observation of *bw*^D-2Rh association is reported by A. Dernburg *et al.* (*Cell*, in the press). □

Received 13 March; accepted 2 May 1996.

1. Heitz, E. *J. b. wiss. Bot.* **69**, 762-818 (1928).
2. Mayfield, J. E. & Ellison, J. R. *Chromosoma* **52**, 37-48 (1975).
3. Lohe, A. R. & Hilliker, A. J. *Curr. Opin. Gen. Dev.* **5**, 746-755 (1995).
4. Henikoff, S., Jackson, J. M. & Talbert, P. B. *Genetics* **140**, 1007-1017 (1995).
5. Henikoff, S. & Dreesen, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6704-6708 (1989).

6. Talbert, P. B., LeCiel, C. D. & Henikoff, S. *Genetics* **136**, 559-571 (1994).
7. Makunin, I. V., Pokholkova, G. V., Zakharkin, S. O., Kholodilov, N. G. & Zhimulev, I. F. *Dokl. Akad. Nauk* **344**, 266-269 (1995).
8. Weiler, K. S. & Wakimoto, B. T. *Rev. Genet.* **29**, 577-605 (1995).
9. Eisenberg, J. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9923-9927 (1990).
10. Kellum, R., Raff, J. W. & Alberts, B. M. *J. Cell Sci.* **108**, 1407-1418 (1995).
11. Locke, J., Kotarski, M. A. & Tartof, K. D. *Genetics* **120**, 181-198 (1988).
12. Vourch, C., Tarascio, D., Boyle, A. L. & Ward, D. C. *Expl. Cell Res.* **205**, 142-151 (1993).
13. Manuelidis, L. *Science* **250**, 1533-1540 (1990).
14. Weiner, B. M. & Kleckner, N. *Cell* **77**, 977-991 (1994).
15. Eberl, D. F., Duyf, B. J. & Hilliker, A. J. *Genetics* **134**, 277-292 (1993).
16. Wakimoto, B. T. & Hearn, M. G. *Genetics* **125**, 141-154 (1990).
17. Swedlow, J. R., Agard, D. A. & Sedat, J. W. *Curr. Opin. Cell Biol.* **5**, 412-416 (1993).
18. Spector, D. L. *Rev. Cell Biol.* **9**, 265-315 (1993).
19. Ashburner, M. *Drosophila: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
20. Smoller, D. A., Petrov, D. & Hartl, D. L. *Chromosoma* **100**, 487-494 (1991).
21. Marshall, W. F., Dernburg, A. F., Harmon, B., Agard, D. A. & Sedat, J. W. *Molec. Biol. Cell* **7**, 825-842 (1996).
22. Sokal, R. R. & Rohlf, F. J. *Biometry* (W. H. Freeman, New York, 1981).
23. Lindsley, D. L. & Zimm, G. G. *The Genome of Drosophila melanogaster* (Academic, San Diego, 1992).

ACKNOWLEDGEMENTS. We thank A. Dernburg and J. Sedat for results and methodologies communicated before publication; J. Henikoff for advice concerning the statistical analysis; Z. Fan for measuring nuclei; P. Goodwin for help with image analysis; and Applied Precision for the loan of the DeltaVision microscope. This work was supported by the Howard Hughes Medical Institute.

CORRESPONDENCE and requests for materials should be addressed to S.H. (e-mail: steveh@howard.fhcre.org).

Regulation of vinculin binding to talin and actin by phosphatidylinositol-4-5-bisphosphate

Andrew P. Gilmore & Keith Burridge

Department of Cell Biology and Anatomy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

VINCULIN, a prominent cytoskeletal protein at cell-substrate adhesions (focal adhesions) and cell-cell adhesions (adherens junctions)¹, interacts with other cytoskeletal proteins, including talin and actin^{2,3}. An intramolecular interaction between the head and tail domains of vinculin masks the binding sites for both proteins^{4,5}. The exposure of cryptic binding sites may be important for promoting focal adhesion assembly. Several agents that induce the formation of focal adhesions act through the GTP-binding protein Rho⁶⁻⁹, which elevates phosphatidylinositol-4,5-bisphosphate (PtdInsP₂) levels by activating phosphatidylinositol-4-phosphate-5-OH kinase (PtdIns-5-OH kinase)¹⁰. PtdInsP₂ regulates several actin-binding proteins, including profilin¹¹, gelsolin¹² and α -actinin¹³, and interacts with vinculin^{14,15}. Here we report that PtdInsP₂ dissociates vinculin's head-tail interaction, unmasking its talin- and actin-binding sites. Microinjection of antibodies against PtdInsP₂ inhibit assembly of stress fibres and focal adhesions.

We examined the effects of PtdInsP₂ on the association of the amino-terminal head domain (relative molecular mass, 90,000; *M*_r 90K) of vinculin with its carboxy-terminal tail. A glutathione *S*-transferase fusion protein (GST/V811-1066) representing the C-terminal tail of vinculin (residues 811-1066) was bound to glutathione agarose beads. Residues 811-1066 contain binding sites for vinculin's head domain, actin, paxillin and acidic phospholipids^{3,5,15,16}. In the absence of lipid, approximately 50% of the vinculin head bound to GST/V811-1066 beads (Fig. 1a, lanes 1 and 2). This association was virtually abolished by 10 μ M ml⁻¹ PtdInsP₂ (Fig. 1a, lanes 3 and 4). Intact vinculin did not interact at all with GST/V811-1066 beads, either in the presence or absence of PtdInsP₂.

Given that PtdInsP₂ inhibited the head-tail interaction, we investigated its effect on vinculin/talin binding. Vinculin was incubated with ¹²⁵I-talin and immunoprecipitated with specific antibodies. PtdInsP₂ increased talin binding to vinculin approximately fourfold (Fig. 1b), but phosphatidylcholine had no effect

on talin binding. Saturation-binding analysis (Fig. 1c) indicated that PtdInsP₂ activated 40% of the vinculin, with a K_d of approximately 30 nM, consistent with previous studies^{2,5}. The 90K head bound talin with a 1:1 stoichiometry and a K_d of around 30 nM. In the absence of PtdInsP₂, there was very little talin binding to intact vinculin, which did not saturate at 600 nM talin. These results

suggest that PtdInsP₂ induces a conformational change in vinculin, exposing the talin-binding site.

We tested whether this conformational change occurs *in vivo* when PtdInsP₂ levels increase following cell adhesion¹⁷. Chicken embryo fibroblasts (CEFs) were plated or kept in suspension for 3 h before immunoprecipitation of vinculin and testing for

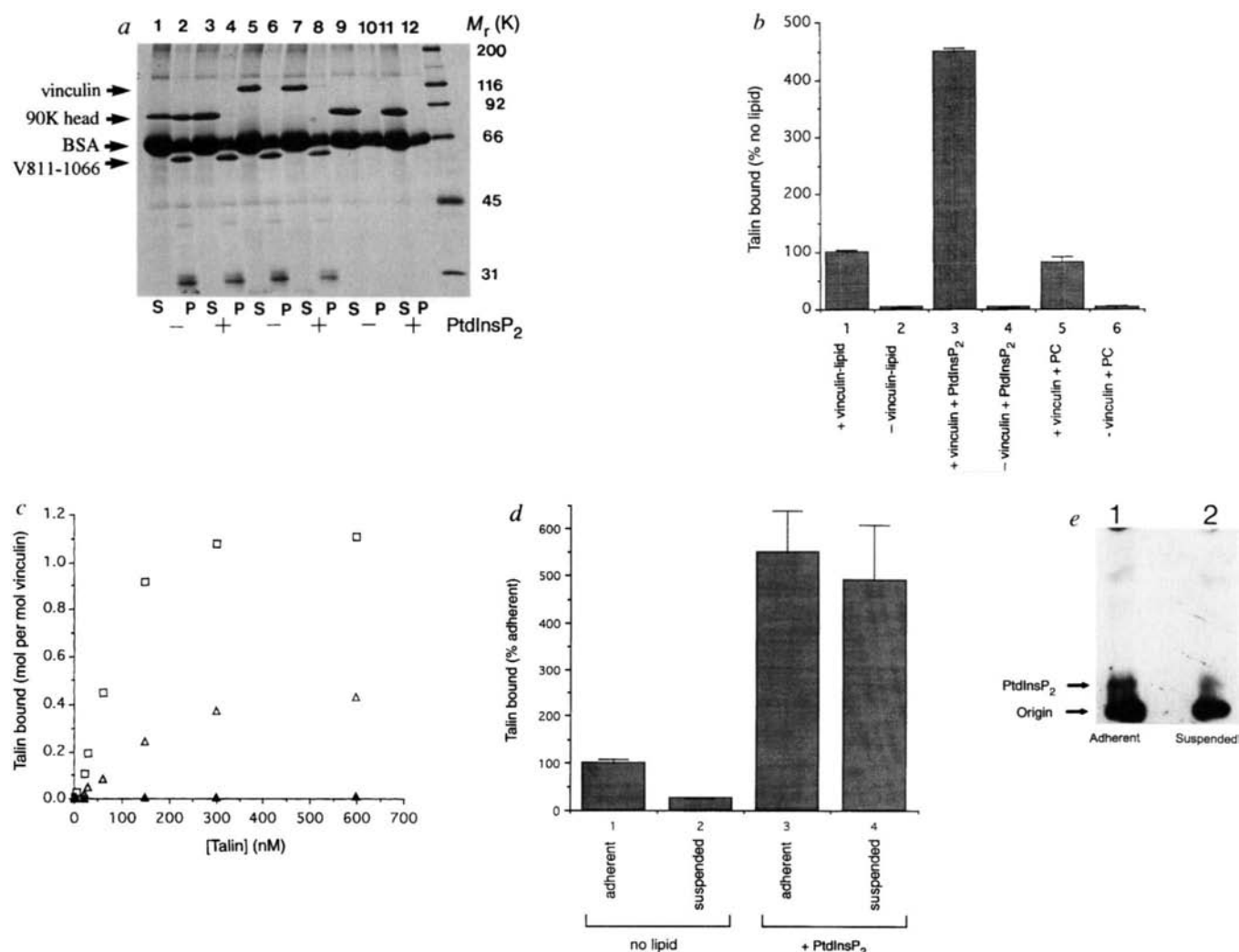


FIG. 1 a, PtdInsP₂ effects on vinculin head-tail interaction. The 90K head or intact vinculin (2 μ M) were incubated with glutathione-agarose beads coated with a vinculin-tail fusion protein (GST/V811-1066). The beads were pelleted, washed and equal amounts of pellet (P) and supernatant (S) resolved by SDS-PAGE and stained with Coomassie blue. Lanes 1-4, 90K with GST/V811-1066; 5-8, intact vinculin with GST/V811-1066; 9-12, 90K with glutathione agarose alone. Lanes 1, 2, 5, 6, 9 and 10, without lipid; 3, 4, 7, 8, 11 and 12, plus PtdInsP₂ (10 μ g ml⁻¹). Note decreased binding of the 90K protein to GST/V811-1066 beads with PtdInsP₂ (lane 4). b, PtdInsP₂ effects on vinculin binding to talin. Intact vinculin (100 nM) was incubated with 10 nM [¹²⁵I]-talin; samples immunoprecipitated with a polyclonal anti-vinculin antibody, washed and counted in an LKB gamma counter. Columns 1 and 2, no lipid; 3 and 4, 10 μ g ml⁻¹ PtdInsP₂; 5 and 6, 10 μ g ml⁻¹ phosphatidylcholine (PC). 1, 3 and 5, with vinculin; 2, 4 and 6, without vinculin. Assays were in triplicate. c, Saturation binding analysis of talin and vinculin with PtdInsP₂. Either the 90K head (open squares) or intact vinculin were incubated with [¹²⁵I]-talin (2-600 nM) with (open triangles) or without (filled triangles) 10 μ g ml⁻¹ PtdInsP₂, and immunoprecipitated as above. d, Analysis of the activation state of vinculin *in vivo*. CEFs, either in suspension for 3 h or plated onto serum-coated culture dishes, were lysed and vinculin immunoprecipitated. Immunoprecipitates incubated with 10 nM [¹²⁵I]-talin, washed and the amount of bound talin determined. 1 and 3, immunoprecipitates from adherent CEFs; 2 and 4, immunoprecipitates from CEFs in suspension; 1, and 2, no added lipid; 3 and 4, 10 μ g ml⁻¹ PtdInsP₂. Talin binding is expressed as the percentage of

that to vinculin from adherent cells without added PtdInsP₂. Background binding in the absence of vinculin has been subtracted. e, Association of PtdInsP₂ with vinculin *in vivo*. CEFs were labelled with [³²P]-orthophosphate and plated or kept in suspension for 3 h. There was no detectable difference in the amount of lipid-incorporated [³²P] between the adherent and suspended samples. Equal amounts of protein were immunoprecipitated with a polyclonal anti-vinculin antibody. Lipids were extracted with chloroform and separated by TLC. Labelled lipids were visualized by autoradiography, and compared with iodine-stained PtdInsP₂ standards. Lane 1, adherent CEFs; 2, CEFs in suspension. No PtdInsP₂ was precipitated by a polyclonal anti-IgG control (not shown).

METHODS. Vinculin, talin, the vinculin 90K V8 protease fragment, and GST/V811-1066 were purified as described^{5,21-24}, and stored in TBS (10 mM Tris-Cl, pH 7.6, 150 mM NaCl, 0.1% β -mercaptoethanol, 0.01% Na₂S₂O₃). Lipids (Sigma) in CHCl₃ were dried under a nitrogen stream and resuspended in TBS, proteins iodinated using 0.5 mCi [¹²⁵I]-Na and iodogen (Pierce), and assays performed in TBS with 0.2% Triton X-100/0.1% bovine serum albumin (BSA), which reduced nonspecific binding. For head-tail binding experiments the supernatant and beads were boiled in sample buffer and subjected to SDS-PAGE²⁵. For *in vivo* PtdInsP₂ assays, identical 60-mm dishes of CEFs were labelled for 18 h with 1 mCi [³²P]-orthophosphate in phosphate-free medium containing 1% fetal bovine serum. Lipids were separated on oxalate-pretreated TLC plates developed with CHCl₃/methanol/ammonium hydroxide. PtdInsP₂ spots were scraped off and quantified by scintillation counting.

^{125}I -talin binding. Vinculin from adherent CEFs bound significantly more talin than that from CEFs in suspension (Fig. 1d, lanes 1 and 2, respectively). Vinculin from both adherent and suspended CEFs exhibited similar increases in talin binding when treated with $10\text{ }\mu\text{g ml}^{-1}$ PtdInsP₂ (Fig. 1d, lanes 3 and 4), suggesting that only a small fraction of vinculin within cells is in the active conformation. We examined the association of PtdInsP₂ with vinculin in adherent and suspended cells labelled with ^{32}P -orthophosphate. Vinculin was immunoprecipitated from equal cell lysates, lipids were extracted, separated by thin-layer chromatography (TLC) and quantified. Between 2.5- and 3-fold more PtdInsP₂ was associated with vinculin immunoprecipitates from adherent CEFs (Fig. 1e, lane 1) compared with CEFs in suspension (Fig. 1e, lane 2).

To investigate the specificity for PtdInsP₂, the effects of other phosphoinositides were examined. The relative effects of phosphatidylinositol (PtdIns), phosphatidylinositol-4-phosphate (PtdInsP), and PtdInsP₂ on vinculin's head-tail interaction are shown in Fig. 2a. PtdInsP₂ inhibited binding of the 90K head to GST/V811-1066-coated beads (Fig. 2a, lanes 7 and 8). Binding was partly inhibited by PtdInsP (Fig. 2a, lanes 5 and 6), where as

PtdIns had no effect (Fig. 2a, lanes 3 and 4). This is consistent with findings that other cytoskeletal proteins regulated by polyphosphoinositides are partly affected by PtdInsP¹¹⁻¹³. Inositol trisphosphate (InsP₃) did not inhibit head-tail binding (not shown). PtdInsP induced an approximately twofold increase in talin binding, compared with a fourfold increase with PtdInsP₂, indicating a correlation between head-tail dissociation and talin binding (Fig. 2b). To show that the effect was dependent on dissociating the head and tail, rather than through effects within vinculin's head or on talin, we examined the binding of ^{125}I -talin to the 90K head in the presence (Fig. 2c, lane 4) or absence (lane 1) of GST/V811-1066; binding was reduced by GST/V811-1066, but restored completely by PtdInsP₂ (lane 6), and partly by PtdInsP (lane 5). In the absence of GST/V811-1066, neither PtdInsP nor PtdInsP₂ affected binding of ^{125}I -talin to the 90K head (lanes 2 and 3).

We examined the role of polyphosphoinositides on vinculin binding to F-actin. In a co-sedimentation assay, intact vinculin bound actin poorly (Fig. 2d, lane 1). PtdInsP₂ increased this threefold, as judged by densitometric scanning of the western blot (Fig. 2d, lane 3). PtdInsP produced an intermediate increase in actin binding (Fig. 2d, lane 2). In the absence of actin, there was

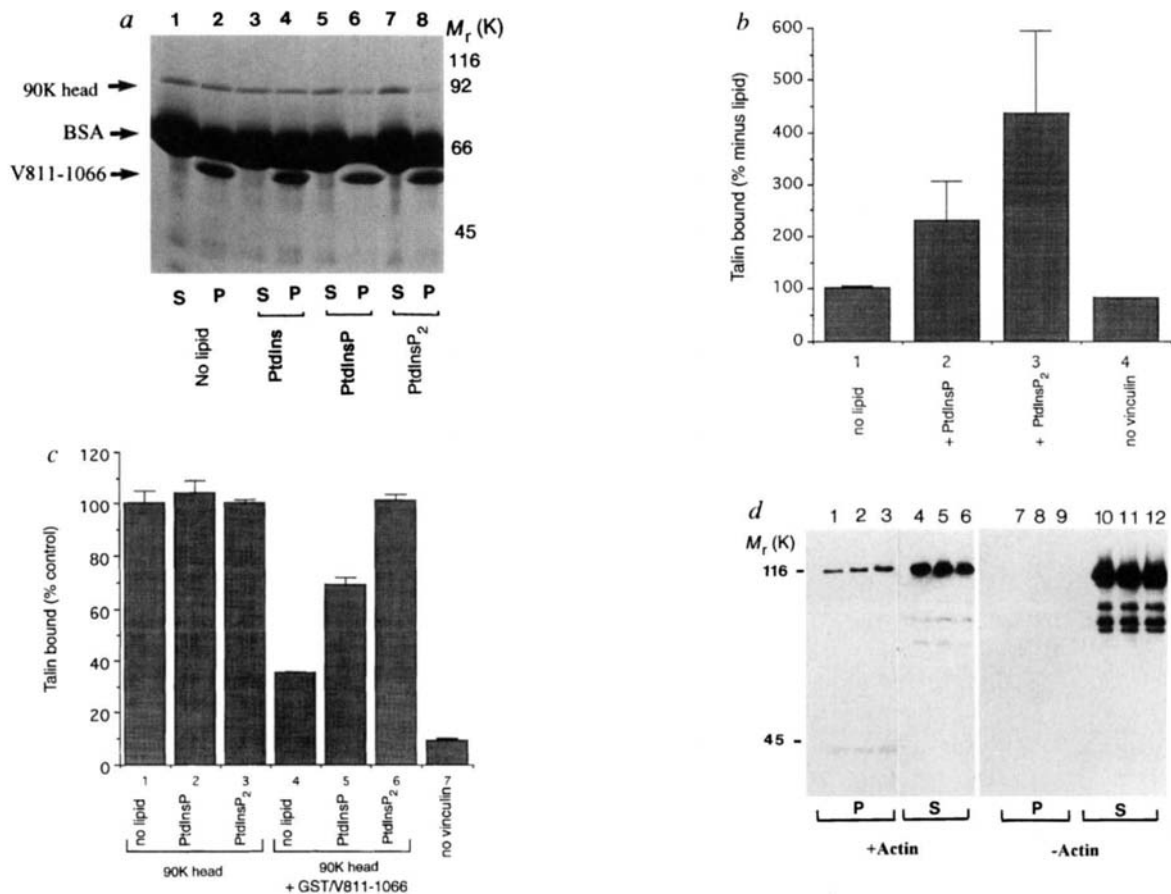


FIG. 2 a, Effects of different phospholipids on the vinculin head-tail interaction; experiment performed as in Fig. 1a. Lanes 1 and 2, no lipid; 3 and 4, PtdIns; 5 and 6, PtdInsP; 7 and 8, PtdInsP₂. Effects of PtdInsP and PtdInsP₂ on vinculin/talin binding. Vinculin was incubated with ^{125}I -talin with or without phospholipids before immunoprecipitating as described above. Column 1, no lipid; 2, PtdInsP; 3, PtdInsP₂; 4, no vinculin. c, Effects of phospholipids on talin binding to the vinculin 90K head with or without GST/V811-1066. Vinculin head (50 nM) was incubated with ^{125}I -talin with or without 100 nM GST/V811-1066. 1, 90K head alone; 2, head plus PtdInsP; 3, head plus PtdInsP₂; 4, head plus GST/V811-1066; 5, head plus GST/V811-1066 and $10\text{ }\mu\text{g ml}^{-1}$ PtdInsP; 6, head plus GST/V811-1066 and $10\text{ }\mu\text{g ml}^{-1}$ PtdInsP₂; 7, ^{125}I -talin without head. d, Effects of PtdInsP and PtdInsP₂ on vinculin's interaction with F-actin. F-actin ($4\text{ }\mu\text{M}$) was polymerized with vinculin (100 nM) with the indicated lipids before

sedimenting. Pellets (P, lanes 1-3 and 7-9) and supernatants (S, lanes 4-6 and 10-12) were separated by SDS-PAGE. Vinculin was detected by western blotting using a polyclonal anti-vinculin antibody: lanes 1-6 with actin, 7-12 without actin. Lanes 1, 4, 7 and 10, no lipid; 2, 5, 8 and 11, $10\text{ }\mu\text{g ml}^{-1}$ PtdInsP; lanes 3, 6, 9 and 12, $10\text{ }\mu\text{g ml}^{-1}$ PtdInsP₂. Lanes 7-12 are overloaded to approximately twice that of lanes 1-6, indicating that no vinculin sedimented without actin.

METHODS. Rabbit skeletal muscle actin was purified as described²⁶, with additional gel filtration. G-actin was stored in 2 mM Tris-Cl, pH 7.6/0.2 mM CaCl₂/0.2 mM ATP/0.1% β -mercaptoethanol. Polymerization was induced by adding 100 mM KCl, 2 mM MgCl₂. Assays were performed in 0.2% Triton X-100 to prevent sedimentation of proteins associated with lipid vesicles. F-actin was sedimented by centrifugation for 30 min at 100,000g in a Beckman TL100 ultracentrifuge.

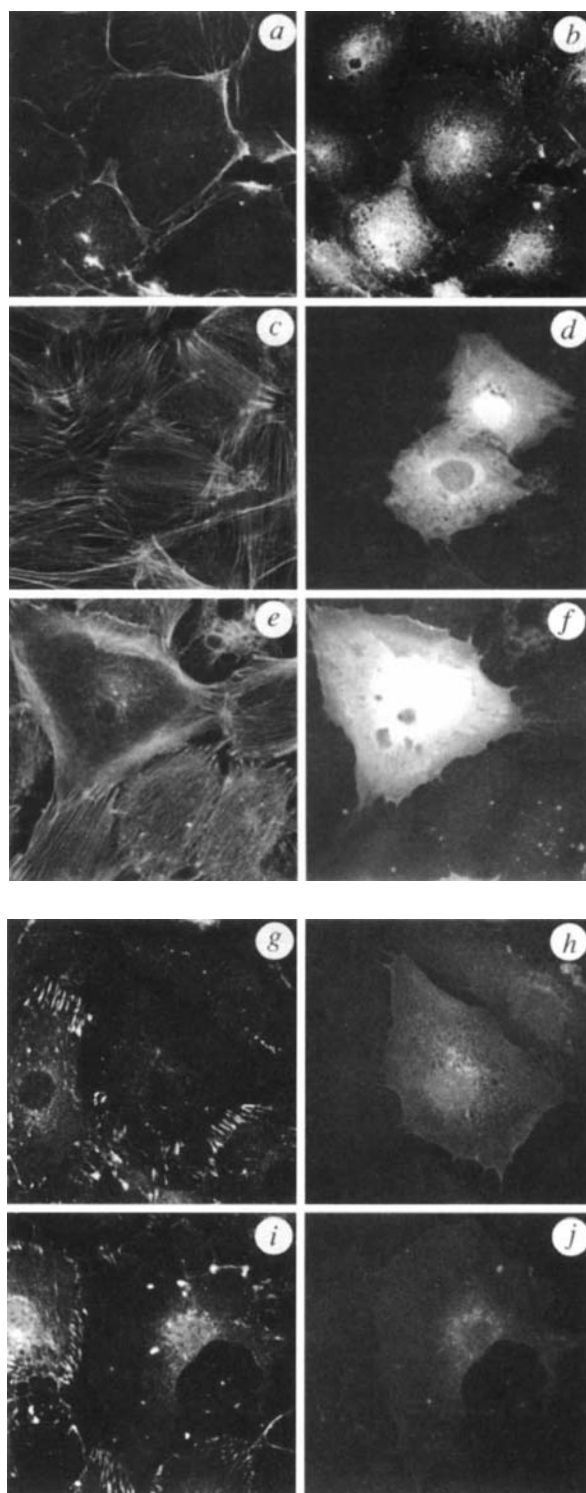


FIG. 3 Sequestering PtdInsP₂ prevents formation of stress fibres and focal adhesions in Balb/c 3T3 cells. Cells were serum starved to induce disassembly of stress fibres and focal adhesions⁶⁻⁸. Starved cells were microinjected with a monoclonal antibody. *a, b*, Starved BALB/c 3T3 cells stained for F-actin (*a*) and vinculin (*b*). *c-j*, Cells stimulated for 20 min with 0.5% fetal bovine serum following antibody injection. Antibodies were co-injected with coumarin-conjugated BSA to identify injected cells (*d, f, h* and *j*). *c, d*, Cells injected with control antibody, stained for F-actin (*c*). *e-j*, Cells were injected with anti-PtdInsP₂ monoclonal antibody and stained for F-actin (*e*), talin (*g*), or vinculin (*i*).

METHODS. Anti-PtdInsP₂ monoclonal antibody was obtained from PerSeptive Diagnostics. Antibodies were dialysed into 75 mM KCl, 10 mM potassium phosphate pH 7.5, 0.1% β -mercaptoethanol. Cells were injected and immunofluorescence performed as described²⁷.

no pelleting of vinculin with either lipid (Fig. 2*d*, lanes 7–9). GST/V811–1066 also binds actin, and is inhibited by the 90K head⁴. We have observed that GST/V811–1066 binding to actin is restored by PtdInsP₂, and partly by PtdInsP (not shown). Paxillin binds the C-terminal tail of vinculin, close to the site that associates with the head domain. We found that paxillin bound vinculin in the absence of lipid, and PtdInsP₂ decreased binding slightly (not shown).

To examine the role of PtdInsP₂ in assembly of focal adhesions, we used serum-starved BALB/c 3T3 cells, which lose stress fibres and focal adhesions (Fig. 3*a, b*). Serum stimulates Rho-dependent formation of focal adhesions and stress fibres⁶⁻⁸. To sequester newly synthesised PtdInsP₂ and prevent it from interacting with cytoskeletal proteins, serum-starved cells were microinjected with a monoclonal antibody against PtdInsP₂¹⁸. Cells injected with anti-PtdInsP₂ did not respond to serum stimulation, and did not form stress fibres or focal adhesions (Fig. 3*e-j*). Injection with a control antibody was not inhibitory (Fig. 3*c, d*). Although the antibody could be affecting several PtdInsP₂-regulated proteins, this result indicates that Rho-induced assembly of focal adhesions requires PtdInsP₂.

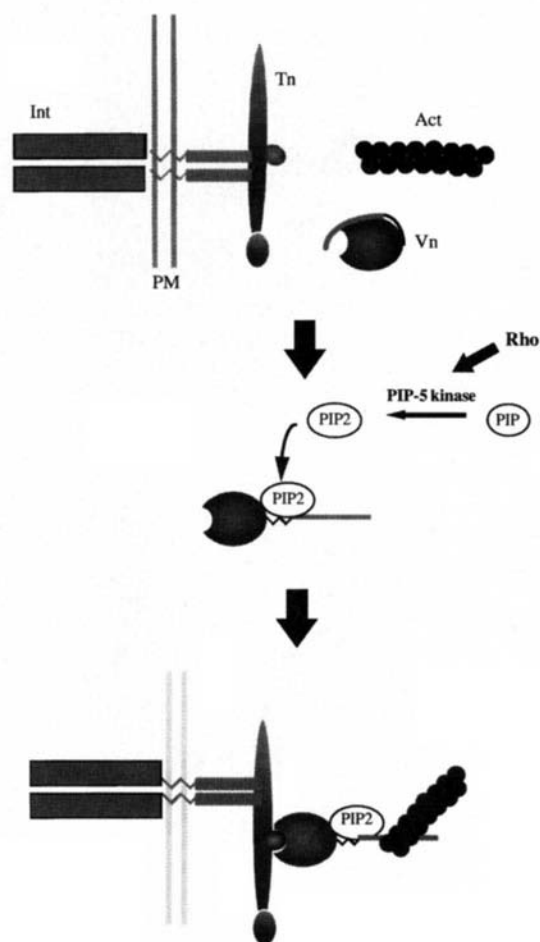


FIG. 4 Model for PtdInsP₂-induced recruitment of focal adhesion proteins. In a resting cell, vinculin (Vn) is in the inactive conformation with the C-terminal tail bound to the head. In this condition it cannot form a link between actin filaments (Act) and proteins such as talin (Tn) bound to integrin (Int)²⁸. The association of proteins with integrin cytoplasmic domains may be induced by binding extracellular matrix proteins²⁹. This interaction can induce an intracellular increase in PtdInsP₂ (PIP2)¹⁷. Activation of Rho through one or more stimuli would lead to activation of PtdIns-5-OH kinase and an increase in PtdInsP₂¹⁰. This may occur in localized regions through association of PtdIns-5-OH kinase with the cytoskeleton³⁰. The association of PtdInsP₂ with vinculin would dissociate the head–tail interaction, exposing binding sites for talin and actin, thereby promoting assembly of focal adhesions.

Assembly of focal adhesions is downstream of Rho, although the pathway has not been elucidated. Tyrosine phosphorylation of several components of focal adhesions (FAK, paxillin and tensin) has been implicated^{19,20}, but the more abundant components, including vinculin, talin and integrins, are not significantly tyrosine-phosphorylated²⁰, suggesting that other signals may regulate their interactions. We suggest a possible model (Fig. 4) in which activation of PtdIns-5-OH kinase by Rho elevates PtdInsP₂. This induces a conformational change in vinculin, promoting its association with other components of focal adhesions and the attachment of microfilaments to the membrane. □

Received 27 November 1995; accepted 15 April 1996.

- Geiger, B., Tokuyasu, K. T., Dutton, A. H. & Singer, S. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4127–4131 (1980).
- Burridge, K. & Mangeat, P. *Nature* **308**, 744–746 (1984).
- Menkel, A. R. et al. *J. Cell Biol.* **126**, 1231–1240 (1994).
- Johnson, R. P. & Craig, S. W. *Nature* **373**, 261–264 (1995).
- Johnson, R. P. & Craig, S. W. *J. Biol. Chem.* **269**, 12611–12619 (1994).
- Ridley, A. J. & Hall, A. *Cell* **70**, 389–399 (1992).
- Barry, S. & Critchley, D. R. *J. Cell Sci.* **107**, 2033–2045 (1994).
- Chrzanowska-Wodnicka, M. & Burridge, K. *J. Cell Sci.* **107**, 3543–3654 (1994).

- Seufferlein, T. & Rozengurt, E. *J. Biol. Chem.* **269**, 9345–9351 (1994).
- Chong, L. D., Traynor-Kaplan, A., Bokoch, G. M. & Schwartz, M. A. *Cell* **79**, 507–513 (1994).
- Lassing, I. & Lindberg, U. *Nature* **314**, 472–474 (1985).
- Janmey, P. A. & Stossel, T. P. *Nature* **325**, 362–364 (1987).
- Fukami, K. et al. *Nature* **359**, 150–152 (1992).
- Fukami, K., Endo, T., Imamura, M. & Takenawa, T. *J. Biol. Chem.* **269**, 1518–1522 (1994).
- Johnson, R. P. & Craig, S. W. *Biochem. biophys. Res. Commun.* **210**, 159–164 (1995).
- Wood, C. K., Turner, C. E., Jackson, P. & Critchley, D. R. *J. Cell Sci.* **107**, 709–717 (1994).
- McNamee, H. M., Ingber, D. E. & Schwartz, M. A. *J. Cell Biol.* **121**, 673–678 (1992).
- Fukami, K. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9057–9061 (1988).
- Burridge, K., Turner, C. E. & Romer, L. H. *J. Cell Biol.* **119**, 893–903 (1992).
- Bockholt, S. M. & Burridge, K. *J. Biol. Chem.* **268**, 14565–14567 (1993).
- Evans, R. R., Robson, R. M. & Stromer, M. H. *J. Biol. Chem.* **259**, 3916–3924 (1984).
- Turner, C. E. & Burridge, K. *Eur. J. Cell Biol.* **49**, 202–206 (1989).
- Groesch, M. E. & Otto, J. *J. Cell Motil. Cytoskel.* **15**, 41–50 (1990).
- Smith, D. B. & Johnson, K. S. *Gene* **67**, 31–40 (1988).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Spudich, J. A. & Watt, S. J. *J. Biol. Chem.* **246**, 4866–4871 (1971).
- Nuckolls, G. H., Romer, L. H. & Burridge, K. *J. Cell Sci.* **102**, 753–762 (1992).
- Horvitz, A., Duggan, K., Buck, C., Beckerle, M. C. & Burridge, K. *Nature* **320**, 531–533 (1986).
- Miyamoto, S., Akiyama, S. K. & Yamada, K. M. *Science* **267**, 883–885 (1995).
- Payraastre, B. et al. *J. Cell Biol.* **115**, 121–128 (1991).

ACKNOWLEDGEMENTS. We thank N. Divecha, L. Romer and M. Schaller for suggestions, and L. Petch, M. Kinch and G. Schneider for comments on the manuscript. This work was supported by an EMBO postdoctoral fellowship (to A.P.G.) and grants from the NIH.

CORRESPONDENCE and requests for materials should be addressed to A.P.G. (e-mail: gilmore@med.unc.edu).

Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA

Arydas Kanopka, Oliver Mühlemann & Göran Akusjärvi

Department of Medical Immunology and Microbiology, BMC, Uppsala University, Uppsala, 751 23 Uppsala, Sweden

THE adenovirus L1 unit¹ represents an example of an alternatively spliced precursor messenger (pre-mRNA) where one 5' splice can be joined to one of two alternative 3' splice sites, producing the 52,55K or the IIIa mRNAs (Fig. 1a). Efficient usage of the distal IIIa 3' splice site requires late viral protein synthesis and is therefore confined to the late phase of virus infection^{2–4}. Here we show that, in extracts from uninfected cells, the classical SR proteins⁵, which are essential splicing factors^{6–7}, inhibit IIIa pre-mRNA splicing by binding to an intronic repressor element and preventing recruitment of the U2 small nuclear ribonucleoprotein particle to the spliceosome. We further show that the viral repressor element has splicing-enhancer activity when appropriately placed in the pre-mRNA. Together, our results demonstrate that SR proteins function as activators or repressors of splicing depending on where on the pre-mRNA they bind.

We have previously shown that in nuclear extracts prepared from uninfected HeLa cells (HeLa-NE) 52,55K splicing is on average 10-fold better than IIIa splicing⁸. To define the *cis*-acting sequence elements responsible for the low splicing activity of the IIIa pre-mRNA in HeLa-NE, we constructed chimaeric pre-mRNAs between IIIa and the efficiently spliced rabbit β -globin pre-mRNA (Fig. 1c). Surprisingly, *in vitro* splicing assays showed that the major *cis*-acting sequence element responsible for the low activity of IIIa is not its weak polypyrimidine tract⁸, but rather a purine-rich intronic repressor element located immediately upstream of the IIIa branch point sequence (Fig. 1b). Thus, transfer of a 110-nucleotide intron element from IIIa to the β -globin pre-mRNA was sufficient to convert it to an inefficiently spliced pre-mRNA (Fig. 1c, transcript 3). Conversely, substituting a 49-nucleotide purine-rich sequence located immediately upstream of the IIIa branch site with β -globin sequences resulted in a superactivation of IIIa splicing compared with the wild-type transcript (Fig. 1c, transcript 5). Collectively, these results indicate that the weak splicing activity of the IIIa pre-mRNA in HeLa-NE

is primarily due to an intronic repressor element located immediately upstream of the IIIa branch site. We have named this 49-nucleotide sequence the IIIa-repressor element (3RE) (Fig. 1b).

SR proteins are a class of essential splicing factors^{5–7} necessary for the early steps in spliceosome formation⁹. There are six classical SR proteins in HeLa cells (Fig. 3a, see later)⁵ but the family is growing^{10,11}. As the SR proteins play a central role in the early recognition of splice sites during spliceosome formation, we tested the activity of purified SR proteins⁵ on IIIa splicing. Addition of SR proteins to splicing-deficient S100 extracts^{8,7} activated 52,55K and β -globin splicing (data not shown), in agreement with their essential role in splicing generic pre-mRNAs⁹. Unexpectedly, none of the individual purified SR proteins was able to activate IIIa splicing in S100 extracts (data not shown). Even more surprisingly, addition of an excess of SR proteins to HeLa-NE almost completely abolished IIIa splicing (Fig. 2a). Furthermore, addition of single HeLa SR proteins showed that each inhibited IIIa splicing, although with different efficiency (Fig. 2a). The SRp30a/b fraction, which consists of a mixture of at least ASF/SF2 and SC35, and SRp55 were most effective in repressing IIIa splicing. To rule out the possibility that the purified SR proteins were contaminated with an unrelated inhibitory factor, we reproduced the SR protein repression of IIIa splicing using a bacterially produced ASF/SF2 protein⁶ (Fig. 3g; see below).

To define the sequence element in the IIIa pre-mRNA necessary for SR protein inhibition, we used chimaeric IIIa/ β -globin pre-mRNAs (Fig. 1c). The results showed that the 3RE (Fig. 1b) was the SR-protein-responsive element (Fig. 2b). Thus, transfer of this genetic element to the β -globin pre-mRNA converted it from an SR-protein-activated to an SR-protein-repressed transcript (Fig. 2b, transcripts 1 and 3). Also, removal of the 3RE completely annulled the inhibitory effect of SR proteins on IIIa splicing (Fig. 2b, transcript 5). Taken together, these experiments show that the 49-nucleotide 3RE contains the minimal sequence element responsible for inhibition by SR proteins.

Why do SR proteins inhibit IIIa splicing? SR proteins have previously been shown to bind RNA and stimulate splicing^{12–19}. Interestingly, the 3RE contains three identical purine-rich sequences which are a good match with the ASF/SF2 consensus RNA-binding sequence²⁰ (Fig. 1b). Crosslinking by ultraviolet radiation of total SR proteins (Fig. 3a) indicated that the SRp30a/b fraction, the most effective inhibitor of IIIa splicing (Fig. 2a), crosslinked most efficiently to the 3RE (Fig. 3e). To a lesser extent, SR proteins also bound to the IIIa branch-site polypyrimidine fragment (Fig. 3c), although binding was strongest to the entire 3' splice site fragment (Fig. 3b). In contrast, SR proteins

FIG. 1 Identification of the IIIa repressor element. *a*, Structure of the adenovirus major late region L1 mRNAs expressed at early and late times of infection¹. The common tripartite leader (1,2,3) is spliced exclusively to the 3' splice site of the 52,55K coding exon (speckled box) at the early stage of infection (early splicing pattern). Late during infection, an additional 3' splice site downstream of the 52,55K coding region is activated, leading to IIIa mRNA production (black box). *b*, Enlargement showing the nucleotide sequence of the IIIa repressor element (3RE), and the IIIa branch point–polypyrimidine tract. Asterisks indicate the two A residues used as branch points in IIIa splicing⁸. *c*, Structure of IIIa⁸ (shaded boxes and thick lines) and β -globin²⁷ (open boxes and thin lines) hybrid pre-mRNAs used to map the 3RE. The black box designates a U1 snRNA binding site that was appended to the 3' end of all transcripts²⁸. The branch point is indicated by an asterisk. The lengths of exons and introns are given above and below each construct, respectively, and the location of the 3RE is shown (striped box). The average splicing efficiency for each construct was derived from the number of experiments shown in parentheses.

METHODS. The IIIa/ β globin hybrid clones (*c*) were constructed by use of appropriately positioned restriction endonuclease cleavage sites and PCR to introduce the additional required site. Detailed plasmid maps are available upon request. ³²P-labelled pre-mRNAs were synthesized and tested for their splicing activity *in vitro* in HeLa-NE as described²².

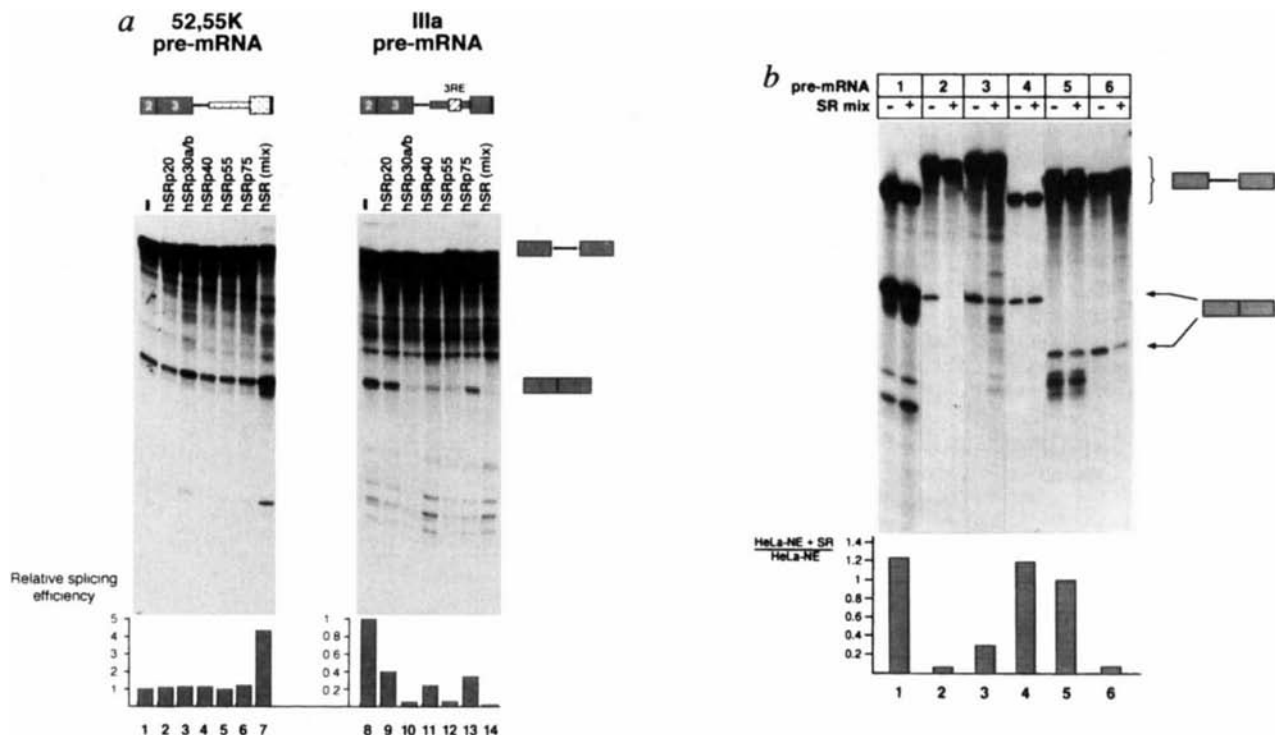
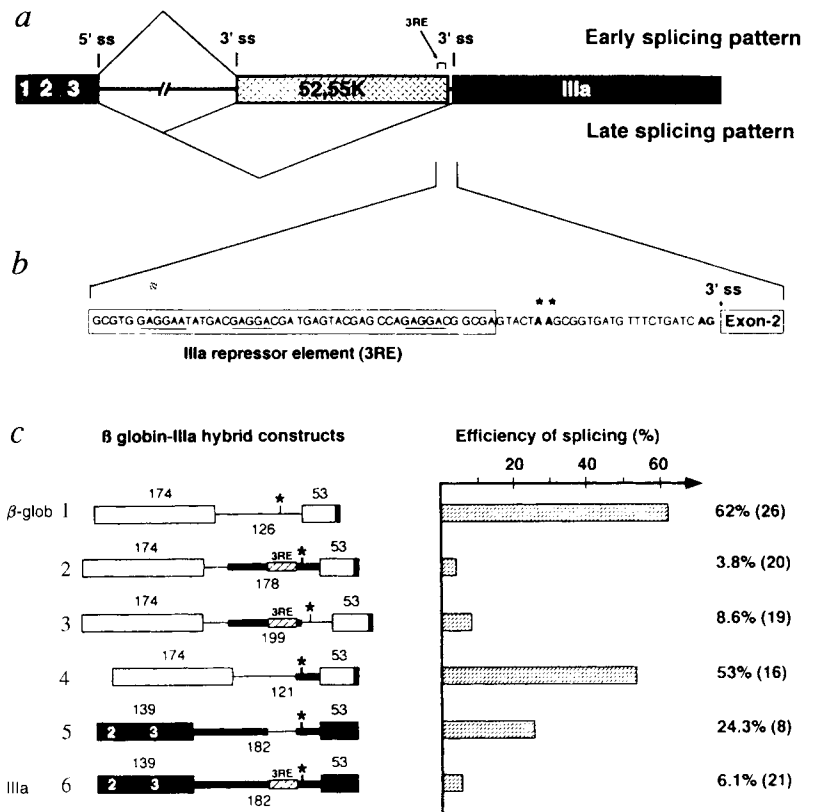


FIG. 2 SR proteins inhibit IIIa pre-mRNA splicing. *a*, Effect of single gel-purified⁵ (lanes 2–6 and 9–13) or the SR protein mixture (lanes 7 and 14) on 52,55K and IIIa pre-mRNA splicing. Bottom, the average splicing efficiencies (from three experiments) compared to the reference (lanes 1 and 8, set as 1) was quantified by PhosphorImager scanning. *b*, Total SR proteins (0.4 μ g, denoted by a + sign) were added to splicing reactions containing the IIIa/ β -globin hybrid pre-mRNAs described in Fig. 1c. The average ratio (from four experiments) of the splicing efficiency + SR proteins/– SR proteins is shown at the bottom. **METHODS.** SR proteins were purified from HeLa spinner cells (*a*) or calf

thymus (*b*)⁵. Note that we have not found any difference in activity between HeLa and calf thymus SR proteins. Splicing reactions contained 0.4–0.5 μ g (this amount of hSRp30a/b efficiently activates 52,55K splicing in S100 extracts^{6,7}; data not shown) of individual SR protein (lanes 2–6 and 9–13) or the SR protein mixture (lanes 7, 14). HeLa-NE extract preparation, substrate RNA synthesis and *in vitro* splicing have been described⁸, except that final splicing reactions contained 2 mM ATP, 12 mM HEPES (pH 7.9), 60 mM KCl, 20 mM creatine phosphate and 0.4 U per μ l RNasin. Splicing products were resolved on 8% denaturing polyacrylamide gels and quantified by PhosphorImager scanning.

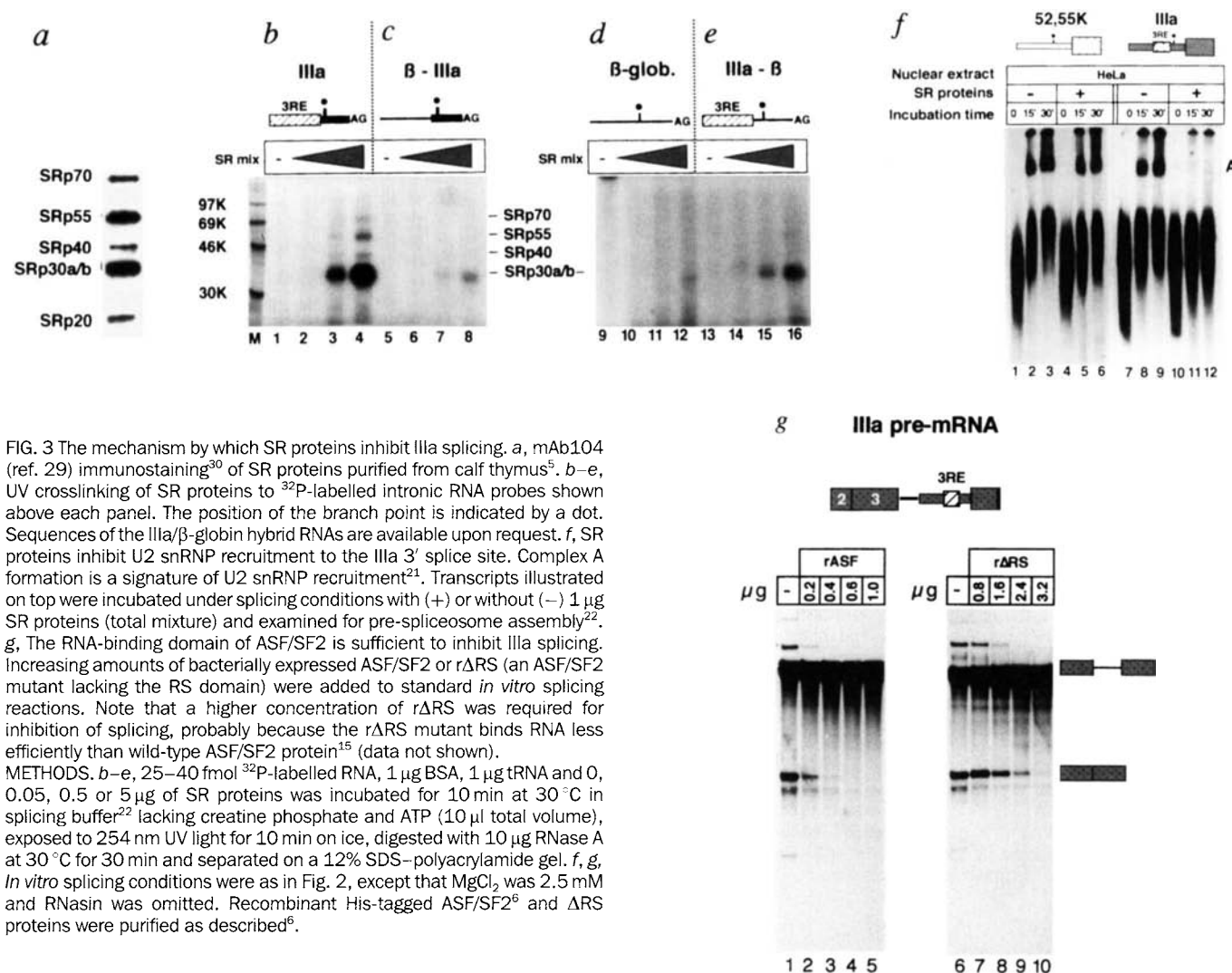


FIG. 3 The mechanism by which SR proteins inhibit IIIa splicing. **a**, mAb104 (ref. 29) immunostaining³⁰ of SR proteins purified from calf thymus⁵. **b–e**, UV crosslinking of SR proteins to ³²P-labelled intronic RNA probes shown above each panel. The position of the branch point is indicated by a dot. Sequences of the IIIa/β-globin hybrid RNAs are available upon request. **f**, SR proteins inhibit U2 snRNP recruitment to the IIIa 3' splice site. Complex A formation is a signature of U2 snRNP recruitment²¹. Transcripts illustrated on top were incubated under splicing conditions with (+) or without (–) 1 μg SR proteins (total mixture) and examined for pre-spliceosome assembly²². **g**, The RNA-binding domain of ASF/SF2 is sufficient to inhibit IIIa splicing. Increasing amounts of bacterially expressed ASF/SF2 or rΔRS (an ASF/SF2 mutant lacking the RS domain) were added to standard *in vitro* splicing reactions. Note that a higher concentration of rΔRS was required for inhibition of splicing, probably because the rΔRS mutant binds RNA less efficiently than wild-type ASF/SF2 protein¹⁵ (data not shown). **METHODS**. **b–e**, 25–40 fmol ³²P-labelled RNA, 1 μg BSA, 1 μg tRNA and 0, 0.05, 0.5 or 5 μg of SR proteins was incubated for 10 min at 30 °C in splicing buffer²² lacking creatine phosphate and ATP (10 μl total volume), exposed to 254 nm UV light for 10 min on ice, digested with 10 μg RNase A at 30 °C for 30 min and separated on a 12% SDS–polyacrylamide gel. **f**, **g**, *In vitro* splicing conditions were as in Fig. 2, except that MgCl₂ was 2.5 mM and RNasin was omitted. Recombinant His-tagged ASF/SF2⁶ and ΔRS proteins were purified as described⁶.

did not bind to the rabbit β-globin 3' splice site (Fig. 3d) or to the 52,55K 3' splice site (data not shown) under our conditions.

To investigate how SR proteins inhibit IIIa splicing, we analysed the effect of SR proteins on spliceosome assembly on the 52,55K and the IIIa pre-mRNAs. Addition of an excess of SR proteins to HeLa-NE efficiently repressed pre-spliceosome and spliceosome formation on the IIIa pre-mRNA, but did not have any adverse effect on spliceosome assembly on the 52,55K pre-mRNA (data not shown). Using 3' splice-site substrate RNAs, we further showed that the SR protein fraction (Fig. 3f) or recombinant ASF/SF2 (data now shown) could block IIIa splicing by inhibiting recruitment of the U2 small nuclear ribonucleoprotein particle (snRNP) to the IIIa 3' splice site, as seen by the absence of A-complex formation²¹. As SR proteins bind to the 3RE (Fig. 1b), they may repress by sterically interfering with recruitment of U2 snRNP to the IIIa branch site. In agreement with this idea, the RNA-binding domain of ASF/SF2 (rARS) is sufficient to repress IIIa splicing (Fig. 3g) and U2 snRNP recruitment to the IIIa 3' branch site (data not shown).

The finding that the 3RE binds SR proteins was surprising because SR protein binding to RNA has previously been shown to enhance the efficiency of splicing^{12,13,15–19}. In previous examples, SR proteins bound to 5' splice sites or to exonic enhancer elements, whereas here they bind to the intron in the IIIa pre-mRNA. To test the possibility that the position of SR-protein binding sites in a pre-mRNA determines whether they function as enhancers or repressors of splicing, we replaced the 3RE in the IIIa pre-mRNA with a splicing enhancer consisting of two consensus ASF/SF2 binding sites²⁰. Addition of recombinant ASF/

SF2 or total SR proteins (data not shown) inhibited IIIa(ASF) splicing (Fig. 4a), demonstrating that an exonic splicing enhancer functions as a splicing repressor when placed at the position of the 3RE. This result also indicated that 3RE might function as a splicing enhancer if appropriately placed in a pre-mRNA, which turns out to be the case (Fig. 4b): moving the 3RE to the second exon of the IIIa pre-mRNA converts it from a splicing repressor to a splicing enhancer.

To test the relevance of SR protein binding to the 3RE for L1 alternative splicing, we constructed a mini 52,55K–IIIa tandem pre-mRNA. As shown in Fig. 4c, splicing of this transcript mimics what is seen during virus infection^{1,22}. Thus, IIIa splicing is significantly increased in nuclear extracts prepared from adenovirus late-infected cells (Ad-NE) compared to HeLa-NE (compare lanes 1 and 6). Interestingly, addition of an excess of SR proteins to Ad-NE or HeLa-NE activated 52,55K splicing while inhibiting IIIa splicing. This result is consistent with the conclusion that the 3RE in this mini construct functions simultaneously as a splicing enhancer in 52,55K splicing and as a splicing repressor in IIIa splicing. Also, our results show that addition of an excess of SR proteins to Ad-NE is sufficient to abolish its enhanced efficiency of IIIa splicing and convert it to an extract with properties similar to nuclear extracts prepared from uninfected cells. Together these results suggest that SR proteins play a major role in the control of L1 alternative splicing during virus infection. The steady-state concentration of SR proteins does not change from early to late virus infection²³, suggesting that the endogenous pool of SR proteins is functionally inactivated at the late phase of the infectious cycle.

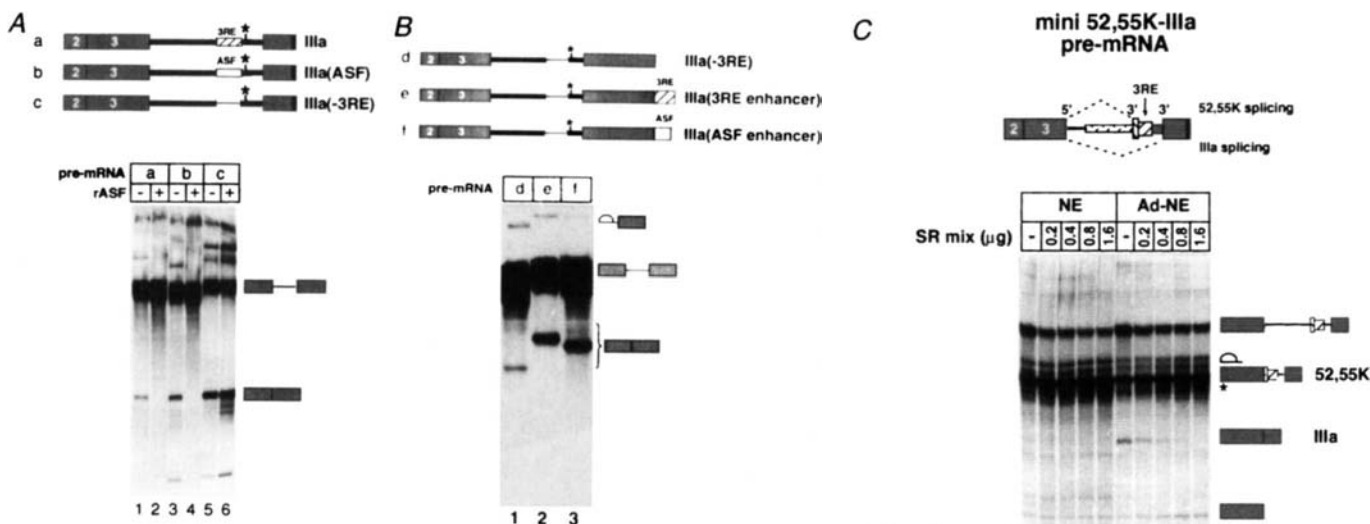


FIG. 4 SR proteins as activators or repressors of splicing. **A**, An ASF/SF2 splicing enhancer functionally substitutes for the 3RE. The transcripts illustrated on top were spliced in HeLa-NE in the presence (+) or absence (–) of 0.5 µg recombinant ASF/SF2. **B**, The 3RE functions as a splicing enhancer when placed in the downstream exon. **C**, SR proteins simultaneously promote 52,55K 3' splice site usage and inhibit IIIa splicing in a tandem 52,55K–IIIa mini construct. Substrate RNA was spliced in uninfected (lanes 1–5) or adenovirus late-infected²² nuclear extract (lanes 6–10) in the presence of increasing amounts of SR proteins. Splicing products were quantified by PhosphorImager scanning and the average values (in per cent) from four experiments are shown at the bottom. Note the difference in scale for 52,55K and IIIa splicing. Asterisks denote degradation products of spliced 52,55K RNA.

METHODS. Substrate RNA synthesis and *in vitro* splicing conditions were as for Fig. 3. The incubation time in **A** and **C** was 2 h and in **B** was 30 min. Transcript **b** was constructed by replacing the 3RE (Fig. 1B) with two consensus ASF/SF2 binding sites²⁰. Transcripts **a** and **c** are identical to 6

and 5 in Fig. 1c, **d** is identical to **c** except for a longer exon 2 (149 nucleotides) and the lack of a U1 snRNA binding site²⁸. The 3RE and the ASF/SF2 binding sites were inserted at the *Mlu*I site in transcript **d** to generate transcripts **e** and **f**, respectively. The mini 52,55K–IIIa construct was generated by deleting a 97-nucleotide fragment located between the 52,55K 3' splice site and the 3RE in pGD52Δ3a²². Sequences and plasmid maps are available upon request.

Most important, our results show that SR proteins act as activators or repressors of splicing depending on whereabouts on the pre-mRNA they bind (Fig. 4). This suggests a major difference with transcriptional enhancers in that splicing enhancers show a clear positional dependence for activity. As SR proteins function as regulators of alternative splicing^{5–7,10,11,24–26}, the mechanism described here, in which SR proteins bind to negative sites in a pre-mRNA and repress splicing by interfering with the recruitment of essential splicing factors to the spliceosome, may be a general way of controlling alternative splicing in mammalian cells. □

22. Kreivi, J.-P. & Akusjärvi, G. *Nucleic Acids Res.* **22**, 332–337 (1994).
23. Himmelsbach, M., Cavaloc, Y., Chebli, K., Stévenin, J. & Gattoni, R. *RNA* **1**, 794–806 (1995).
24. Mayeda, A. & Krainer, A. R. *Cell* **68**, 365–375 (1992).
25. Fu, X. D., Mayeda, A., Maniatis, T. & Krainer, A. R. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11224–11228 (1992).
26. Zahler, A. M., Neugebauer, K. M., Lane, W. S. & Roth, M. B. *Science* **26**, 219–222 (1993).
27. Hornig, H., Aebi, M. & Weissmann, C. *Nature* **324**, 589–591 (1986).
28. Kreivi, J.-P., Zerivitz, K. & Akusjärvi, G. *Nucleic Acids Res.* **19**, 6956 (1991).
29. Roth, M. B., Murphy, C. & Gall, J. G. *J. Cell Biol.* **111**, 2217–2223 (1990).
30. Ausubel, F. M. et al. in *Current Protocols in Molecular Biology* 10.8.1–10.8.17 (Wiley, New York, 1995).

ACKNOWLEDGEMENTS. A.K. and O.M. contributed equally to this work. We thank J. Manley and J. Harper for the wild-type ASF/SF2 plasmid and invaluable help in establishing the S100 splicing assay, and J.-P. Kreivi for the rΔASF mutant and much help at the early stages of this project. This work was supported by the Swedish Cancer Society.

Received 9 February; accepted 17 April 1996.

1. Imperiale, M., Akusjärvi, G. & Leppard, K. in *Current Topics in Microbiology* (eds Doerfler, W. & Böhm, P.) 139–171 (Verlag Chemie, Berlin & Heidelberg, 1995).
2. Akusjärvi, G. & Persson, H. *Nature* **292**, 420–426 (1981).
3. Nevins, J. R. & Wilson, M. C. *Nature* **290**, 113–118 (1981).
4. Svensson, C. & Akusjärvi, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4690–4694 (1986).
5. Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. *Genes Dev.* **6**, 837–847 (1992).
6. Ge, H., Zuo, P. & Manley, J. L. *Cell* **66**, 373–382 (1991).
7. Krainer, A. R., Mayeda, A., Kozak, D. & Binns, G. *Cell* **66**, 383–394 (1991).
8. Kreivi, J.-P., Zerivitz, K. & Akusjärvi, G. *Nucleic Acids Res.* **19**, 2379–2386 (1991).
9. Fu, X.-D. *RNA* **1**, 663–680 (1995).
10. Sreteron, G. R. et al. *EMBO J.* **14**, 4336–4349 (1995).
11. Cavaloc, Y., Popielarz, M., Fuchs, J.-P., Gattoni, R. & Stévenin, J. *EMBO J.* **13**, 2639–2649 (1994).
12. Staknis, D. & Reed, R. *Molec. cell. Biol.* **14**, 7670–7682 (1994).
13. Kohlt, J. D. et al. *Nature* **368**, 119–124 (1994).
14. Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A. & Krainer, A. R. *EMBO J.* **12**, 3607–3617 (1993).
15. Zuo, P. & Manley, J. L. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3363–3367 (1994).
16. Tanaka, K., Watakabe, A. & Shimura, Y. *Molec. cell. Biol.* **14**, 1347–1354 (1994).
17. Tian, M. & Maniatis, T. *Cell* **74**, 105–114 (1993).
18. Lavigne, A., La, B. H., Kornblith, A. R. & Chabot, B. *Genes Dev.* **7**, 2405–2417 (1993).
19. Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R. & Rottman, F. M. *Genes Dev.* **7**, 2598–2608 (1993).
20. Tacke, R. & Manley, J. L. *EMBO J.* **14**, 3540–3551 (1995).
21. Moore, M., Query, C. C. & Sharp, P. A. in *The RNA World* (eds Gesteland, R. F. & Atkins, J. F.) 303–357 (Cold Spring Harbor, New York, 1993).

CORRESPONDENCE and requests for materials should be addressed to G.A. (e-mail: goran.akusjarvi@imim.uu.se).

CORRECTION

Histology of the first fish

M. P. Smith, I. J. Sansom & J. E. Repetski

Nature **380**, 702–704 (1996)

THE lettering used to indicate features on the scanning electron micrograph shown in Fig. 2b unfortunately did not reproduce properly. Labelling of the arrows (top to bottom) should have read tdn, dt, gl; these abbreviations are defined in the legend. □